

AD-A117 722

ARIZONA UNIV TUCSON DIGITAL IMAGE ANALYSIS LAB
OPTICAL COMPUTATIONS FOR IMAGE BANDWIDTH COMPRESSION.(U)
MAY 82 B R HUNT, R SCHOWENGERDT

F/6 9/2

UNCLASSIFIED

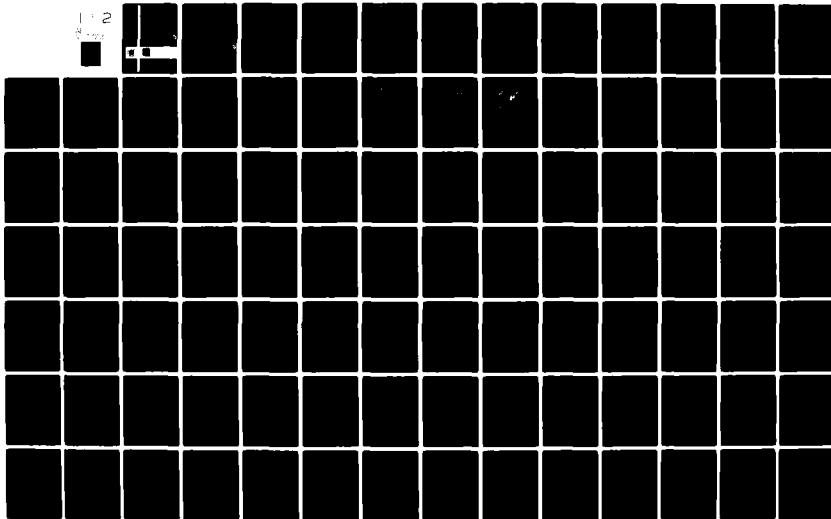
DIAL/EE-81-005

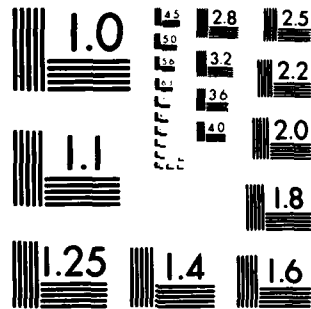
AFOSR-TR-82-05A0

AFOSR-81-0170

NL

1 2





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AFOSR-TR- 82 - 0580 .

2

Final Technical Report

Grant No. AFOSR-81-0170

OPTICAL COMPUTATIONS FOR IMAGE
BANDWIDTH COMPRESSION

Prepared for

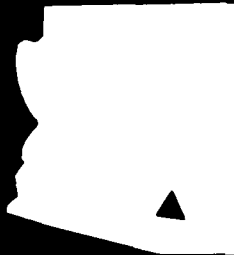
Air Force Office of Scientific Research
Electronic and Solid-State Sciences Division
Bolling Air Force Base
Washington, D. C. 20332

Prepared by

B. R. Hunt
R. A. Schowengerdt

15 MAY 1982

DTIC
ELECTE
AUG 3 1982
KH



DTIC FILE COPY

Approved for public release;
distribution unlimited.

ENGINEERING EXPERIMENT STATION
COLLEGE OF ENGINEERING

THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721

82 08 03 107

EE/DIAL-82-005

OPTICAL COMPUTATIONS FOR IMAGE
BANDWIDTH COMPRESSION

by

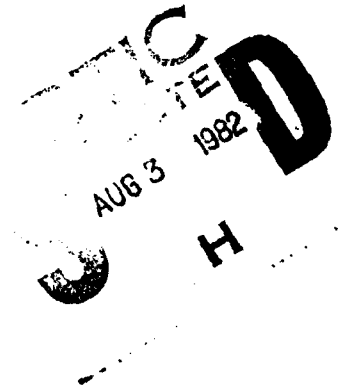
B. R. HUNT

May 15, 1982

Final Technical Report

Research sponsored by the Air Force Office of
Scientific Research under Grant No. AFOSR-81-0170.

Digital Image Analysis Laboratory
Department of Electrical Engineering
University of Arizona
Tucson, Arizona 85721



AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DTIC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12.
Distribution is unlimited.
MATTHEW J. KENPER
Chief, Technical Information Division

I

Table of Contents

Page

(I)	Report Documentation Page	1
(II)	Disclaimer	2
(III)	Introduction	3
(IV)	List of Publications Resulting from AFOSR Support	4
(V)	Personnel	6
Appendix I	Optical Sub-Sampling for Multispectral Image Data Compression	7
	Image Sampling, Reconstruction, and the Effect of Sample-Scene Phasing	19
Appendix II	Optical/Digital Hybrid System Architectures for Interframe (Temporal) Image Data Compression	60
	A Temporally Adaptive Hybrid Optical/ Digital Interframe Compression Scheme	91
Appendix III	Radiometric and Spatial Transformations for Image Data Compression	100
	Adaptive Data Compression for Generating Stationary Statistical Image Models	130



Accession For	
NTIS GSA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

1

(I) REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 82-0580	2. GOVT ACCESSION NO. AD-A117 722	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Optical Computations For Image Bandwidth Compression		5. TYPE OF REPORT & PERIOD COVERED Final Report March 15, 1981-March 14, 1982
6. AUTHOR(s) B. R. Hunt R. Schowengerdt		7. PERFORMING ORG. REPORT NUMBER DIAL/EE 81-005
8. PERFORMING ORGANIZATION NAME AND ADDRESS Digital Image Analysis Laboratory Department of Electrical Engineering University of Arizona, Tucson, Arizona 85721		9. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0170
10. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NE Bldg. 410, Bolling Air Force Base District of Columbia 20332		11. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2305/B1
12. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. REPORT DATE May 15, 1982
14. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15. NUMBER OF PAGES 137
16. DISTRIBUTION STATEMENT (of this Report)		17. SECURITY CLASS. (of this report) Unclassified
17. DISTRIBUTION STATEMENT (of this Report)		18. DECLASSIFICATION/DOWNGRADING SCHEDULE
19. SUPPLEMENTARY NOTES		
20. KEY WORDS (Continue on reverse side if necessary and identify by block number) Optical Computations Electro-Optical Computations Image Bandwidth Compression Spline/Polynomial Interpolation Human Visual Models		
21. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report summarizes the previous year's research in optical processing and optical/digital processing methods for image data compression. The following topics are discussed; (1) Radiometric and spatial transformations to achieve adaptive image data compression;		

UNCLASSIFIED

UNCLASSIFIED

- I
- (2) Spatial optical subsampling and spectral/spatial reconstruction for image data compression;
 - (3) Temporal interframe data compression.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(II.) Disclaimer

The findings in this report are solely those of the author and are not to be interpreted as the official position of the Air Force of Scientific Research or the U. S. Government.

This document constitutes the report on Grant AFOSR-81-0170 for the period from March 15, 1981 to March 14, 1982. The format under which the report is organized is described as follows. There were three major and separate research tasks carried out during the past year under the sponsorship of Grant AFOSR-81-0170. These separate research task areas were: optical sub-sampling for multispectral image data compression; optical/digital hybrid system architectures for interframe (temporal) image data compression; the usage of radiometric and spatial transformations for adaptive image data compression. Since each of these research task areas is independent of the others, we have chosen a report format which emphasizes the independence and makes it possible to concentrate upon reading about the activity in a particular topic area with no involvement or conflict with topical material from another area. Thus, the major research results are presented in Appendices I, II, and III, three appendices that can be separated from the rest of the report, if so desired. As such, each appendix could be considered a "mini-report". Each appendix has its own introductory section which summarizes the research objectives of the research reported in that appendix.

LIST OF PUBLICATIONS RESULTING FROM AFOSR SUPPORT

- (1) B.R. Hunt, "Some remaining problems in nonlinear image restoration", Proc. Symp. Image Science Mathematics, Naval Postgrad. School, Monterey, Calif., Nov. 10-12, 1976.
- (2) B.R. Hunt, "An optical analogy to DPCM digital image data compression", Proc. SPIE, Vol. 118, August, 1977.
- (3) D.G. McCaughey, "An image coding algorithm using spline functions", Proc. SPIE, Vol. 149, August, 1978.
- (4) D.G. McCaughey and H.C. Andrews, "Image approximation by variable knot bicubic splines", to be published, IEEE Trans. Pattern Analysis and Machine Intelligence.
- (5) B.R. Hunt, "Optical computations for image bandwidth compression: analysis and simulation", Appl. Optics, vol. 17, pp. 2944-2951, 1978.
- (6) D.J. Granrath and B.R. Hunt, "Signal detection trade-off analysis of optical vs. digital Fourier transform computers", Appl. Optics, vol. 18, pp. 36-43, 1979.
- (7) D.J. Granrath and B.R. Hunt, "A two-channel model of image processing in the human retina", Proc. SPIE, vol. 199, August, 1979.
- (8) B.R. Hunt, "Nonstationary statistical image models", Computer Graphics and Image Processing, vol. 12, pp. 173-186, 1980.
- (9) D.G. McCaughey, "Variable resolution hybrid-coding Techniques", ASSP Workshop on Two-Dimensional Signal Processing, Berkeley, 1979.
- (10) D.G. McCaughey and H.C. Andrews, "The continuous-discrete model: least-squares inverses and singular function expansions", accepted for publication, IEEE Trans. Info Theory.

- (11) B. R. Hunt and S. D. Cabrera, "Optical implementation of a spatially adaptive image data compression scheme", Proc. SPIE, vol. 232, Washington, 1980.
- (12) B. R. Hunt, and H. Ito, "A hybrid optical/digital interframe image data compression scheme", Proc. SPIE, vol. 249, San Diego, 1980
- (13) B. R. Hunt and S. D. Cabrera, "Optical computations for a spatially adaptive image data compression system", Optical Engineering, vol. 20, pp. 616 - 620, 1981.
- (14) H. Ito & B. R. Hunt, "A temporally adaptive hybrid optical/digital interframe compression scheme", Proc. SPIE, vol. 292, San Diego, 1981.
- (15) W. R. Stevens & B. R. Hunt, "Software pipelines in image processing", accepted for publication, Computer Graphics and Image Processing.
- (16) R. Strickland, "Transforming images into statistically stationary behavior", In Preparation.
- (17) R. N. Strickland, "Adaptive data compression by transformations for generating stationary statistical image models", accepted for Proceedings of the York Conference on Image Processing, York, U. K., 1982.

The following is a list of personnel who received support from Grant AFOSR-81-0170 during the period from March 15, 1981 to March 14, 1982:

Professor B. R. Hunt: Principal Investigator

Assistant Professor R. N. Strickland: Department of
Electrical Engineering

Assistant Professor R. Schowengerdt: Remote Sensing and
Department of Electrical Engineering

Mr. Richard Stevens: Ph.D candidate in Department of Systems
Engineering

Mr. Hiroyasu Ito: Ph.D candidate in Department of Systems
Engineering

Mr. Robert Gray: Ph.D candidate in Optical Sciences Center

Mr. Warren Smith: first year Ph.D student in Optical
Sciences Center

Appendix I

Optical Sub-Sampling for Multispectral Image Data Compression

Optical Sub-Sampling for Multispectral Image Data Compression

Research Objectives

Conventional methods in multi-spectral image data compression require de-correlation transformations, e.g., the Karhunen-Loeve, which are very costly in computing. The objective of this task is to demonstrate that substantial data compression (e.g., 4:1) can be achieved with little sacrifice in quality. The method is to under sample each image array of a multispectral system, but with the sample array of each spectral band staggered or offset so that sample pixels missing in a given band are sampled in other bands. The use of the combined spatial and spectral correlation within and between bands is used to estimate missing pixels and reconstruct imagery of higher quality than that obtained solely from under-sampled data.

Appendix Contents:

This appendix consists of a summary of the research in this area as directly evaluated to date, plus a paper on general problems in image sampling which is of relevance to the sub-sampling which is employed in this method.

This section of the report will review progress on the project to achieve data compression of multispectral imagery by staggered array sensor undersampling and subsequent reconstruction of the original scene utilizing interband redundancy of edge information. The technique assumes that features possessing high spatial frequencies are similar in all spectral bands of a scene. This will be true for shadow borders; however, spectrally dependent edges such as the junction of materially different terrain regions may be less spectrally correlated.

Prototype solid-state cameras have been constructed which use similar staggered color (RGB) arrays [1], [2], [3]. The device in [1] comprises a set of adjoining linear CCD's, each array being sensitive to only one spectral band. The arrays are deployed in alternating color sensitivity, with elements of each strip offset from the elements of neighboring strips. Dye-deposition color filter arrays (CFA's) have been bonded to sensor array chips [2], [3]. The CFA's are rectilinear masks with repeating pattern $\begin{smallmatrix} R & G \\ G & B \end{smallmatrix}$; the mask is thus staggered by spectral element but without sensor gaps as was the case in [1]. Configurations up to 484×384 elements with $34 \mu\text{m} \times 20 \mu\text{m}$ element size have been achieved [3]. However, little attempt has been made to estimate imagery in unsampled bands from data in a sampled band: in [2], edge information in the high resolution G band was added directly to the low resolution R and B bands.

The sampling scheme considered in the present research consists of four bands of a multispectral sensor which are equally

undersampled, although at full resolution, at staggered intervals in a scene. The result is a single composite image, a mosaic of the band samples, having identical spatial dimensions to the original scene. The four spectral estimates of the scene are then to be reconstructed in full detail from this single image. The spatial arrangement of pixels in a four-band image, bands I - IV, is

I	II	I	II	
III	IV	III	IV	
I	II	I	II	...
III	IV	III	IV	
	:			
	:			
	:			

A reconstruction window five samples square is passed across the mosaic image; at each window location, the center sample, plus estimates of the other three bands formed from the window content, are written to the corresponding spatial coordinates in four output images.

The contents of the window may be described by the array:

X	H	S	H	X
V	D	V	D	V
S	H	S	H	S
V	D	V	D	V
X	H	S	H	X

where

S is the band of the central sample,

H is the band of the horizontally nearest-neighbor sample,

V is the band of the vertically nearest-neighbor sample, D is the band of the diagonally nearest-neighbor sample, and X positions are ignored.

In each window position the value of S is given, and the values of H, V, and D must be estimated from the contents of the window. If E is the general band to be estimated, (i.e., H, V, or D) then the values S and E may be decomposed into local means and differences as:

$$\begin{aligned} S &= \bar{S} + \Delta S \\ E &= \bar{E} + \Delta E. \end{aligned} \quad (1)$$

Let us form the estimate $\hat{E} = \bar{E} + \alpha \Delta S$, i.e., we wish to estimate ΔE from ΔS . A MMSE criterion such that $\langle (E - \hat{E})^2 \rangle$ a minimum will be assumed. The band images of E and S will be treated as weakly stationary random variables over the extent of the estimation window, so that the ensemble averages are replaced by spatial averages. The coupling parameter α resulting from the MMSE optimization is

$$\alpha = \rho_{ES}(\sigma_E/\sigma_S), \quad (2)$$

where σ_E and σ_S are the standard deviations of the E and S bands, respectively, and ρ_{ES} is the correlation coefficient between E and S, defined by

$$\rho_{ES} = \frac{\langle \Delta E \Delta S \rangle}{\sigma_E \sigma_S}, \quad -1 \leq \rho_{ES} \leq 1. \quad (3)$$

Note that negative values of ρ account for contrast reversals between bands, and that for $\rho = 0$ (uncorrelated E and S bands) the band estimate is the local mean. If α is correctly chosen according to (2), then the residual mean square error ϵ^2 depends

on ρ_{ES} :

$$\epsilon^2 = \sigma_E^2(1 - \rho_{ES}^2) . \quad (4)$$

When $\rho_{ES} = \pm 1$, perfect reconstruction is possible; when E and S are uncorrelated (and when $\alpha = 0$), then ϵ^2 equals the variance of the estimated band. The typical situation will fall between these extremes.

The estimation of the correlation coefficient ρ_{ES} at each window location is non-trivial. If the E and S band values of each window element were a priori known (in which case no reconstruction would be needed), the expectation $\langle \Delta E \Delta S \rangle$ could be formed from the average product of the E and S differences at each element due to the assumption of local stationarity. In the absence of such full information, an initial approximation of ρ_{ES} may be obtained by averaging the product of interpolated values at each window element. The accuracy of the ρ_{ES} approximation will then depend on the within-band spatial correlations of both E and S.

Simulation

The above algorithm has been used in the reconstruction of a demonstration image. The upper left photograph in figure 1 shows the original green band of a tricolor (RGB) 256x256 8-bit digitized image. In place of four distinct bands, the mosaic will be formed from a sampling array $\begin{smallmatrix} R & G \\ B & R \end{smallmatrix}$, with the R red band in both positions I and IV. The three associated maps ($G-R_I$, $G-R_{IV}$, $G-B$) are included in figure 1 clockwise from the top. A

closeup of the original is given in figure 2, and figure 3 displays the mosaic sensor image.

It is often true that some pairs of bands are more spectrally correlated than others. This knowledge can be used in forming the final band estimate. For example, in the demonstration image bands G-B and bands R_I-R_{IV} are such pairs. A post-reconstruction filter can be devised that replaces an estimate by its within-band nearest neighbors average if the local $|\rho_{SE}|$ is below a threshold and if the local S and E bands are not high-correlation pairs. The sample array symmetry then guarantees that the nearest-neighbor within-band estimates were formed from high-correlation pairs and tend to be more reliable estimates. Thus the best surrounding reconstruction estimates will supplant a poor estimate.

Figure 4 shows the result of reconstruction of the mosaic-sampled image of figure 2, after the described post-filtering with a $|\rho_{SE}|$ threshold of 0.5. Approximations of ρ_{SE} were generated from bilinear-interpolated versions of the mosaic. For comparison, a direct bilinear-interpolation of the mosaic sampled original displays considerable blurring (figure 5). Cubic spline and sinc interpolation results are similarly degraded.

The root normalized mean square error (RNMSE) between the original and estimate of each band is given in the following table:

Demonstration Image - RNMSE	R-I	G-II	B-III	R-IV
Reconstruction,				
"true ρ " maps	.02917	.06127	.06590	.02869
Post-filtered version of				
above	.02822	.05960	.06464	.02773
Reconstruction,				
Interpolated ρ maps	.02993	.06235	.06941	.02931
Post-filtered version				
of above	.02875	.06161	.06842	.02832
Bilinear interpolation	.03247	.07286	.07605	.03206

The correlation-reconstructions are seen to be more accurate than the bilinear interpolation. The replacement post-filtering, with a $|\rho_{SE}|$ threshold of 0.5, makes a small improvement in the estimation of all bands.

Summary

Test results have indicated that interband correlation is useable in a simple multispectral data compression scheme whose only coding is the spatial configuration of the sensor array. The 4:1 maximum compression ratio for the four-band case does not take into account any further compression available by the usual single-band methods, e.g., DPCM. Future research will consider the effect of the sampling scheme on single-band methods. The performance of various reconstruction-window configurations, the treatment of nonstationarity within the window, and simulations on a more varied selection of imagery will also be examined.



Figure 1. Green Band Original with Correlation Coefficient Maps



Figure 2. Detail of Original, Single-Band

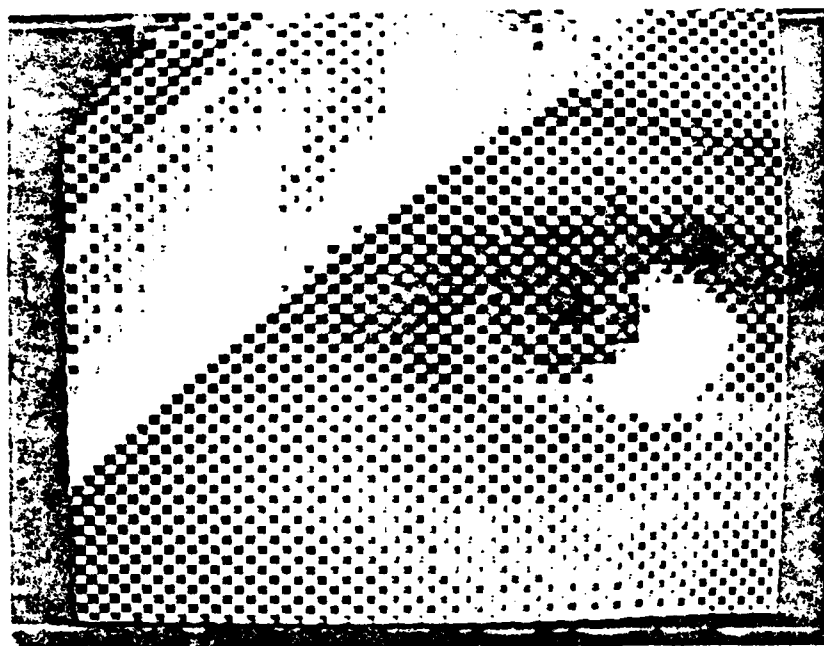


Figure 3. Mosaic Sampling of Four Bands



Figure 4. Reconstructed and Post-Filtered Estimate of Single Band



Figure 5. Bilinear Interpolation Estimate of Single Sand

REFERENCES

- [1] S. Ochi, S. Yamanaka, Y. Kanoh, and T. Nishimura, "A Device Structure and Spatial Spectrum for Checker-Pattern CCD Color Camera", IEEE Trans. Electron Devices, Vol. ED - 25, pp. 261 - 266, 1978.
- [2] P. Dillon, D. Lewis, and F. Kasper, "Color Imaging System Using a Single CCD Area Array", IEEE Trans. Electron Devices, Vol. ED - 25, pp. 102 - 107, 1978.
- [3] N. Koike, I. Takemoto, R. Satoh, S. Hanamura, S. Nagahara, and M. Kubo, "MOS Area Sensor" Part I- Design Consideration and Performance of an n-p-n Structure 484x484 Element Color MOS Imager", IEEE Trans. Electron Devices, Vol. ED - 27, pp. 1676 - 1681, 1980.

IMAGE SAMPLING, RECONSTRUCTION, AND THE
EFFECT OF SAMPLE-SCENE PHASING

Stephen K. Park*

Robert A. Schowengerdt**

Submitted to Applied Optics

March 1982

* NASA, Langley Research Center, Hampton, VA 23665

** University of Arizona, Office of Arid Lands Studies
and Electrical Engineering Department, Tucson, AZ 85721.
Support from Grant AFOSR-81-0170 for this research is
gratefully acknowledged.

Abstract

This paper is a one-dimensional analysis of the degradation caused by image sampling and interpolative reconstruction. The analysis includes the sample-scene phase as an explicit, random parameter and provides a complete characterization of this image degradation as the sum of two terms; one term accounts for the mean effect of undersampling (aliasing) and non-ideal reconstruction averaged over all sample-scene phases; the other term accounts for variations about this mean. The results of this paper have application to the design and performance analysis of image scanning, sampling and reconstruction systems.

Introduction

In 1934 Mertz and Gray¹ published their pioneering paper on the performance of a particular line-scan imaging system, television. Over the next 20 years this work was extended considerably by Schade² and others^{3,4}. More recently a diversity of line-scan imaging systems have evolved from the application of digital and electro-optical technology in such areas as space exploration, remote sensing, reprographics, and medicine. These more recent scanning systems typically employ electronic sampling and digital (or optical) image reconstruction and their performance has been analyzed in numerous papers and texts, for example references 5 through 12. Although various effects have been considered in these performance analyses, one which is commonly ignored is that associated with the phase uncertainty of the scene relative to the sampling grid.

It is well known that for practical systems, image sampling and reconstruction inevitably produce some degradation in the reconstructed image¹²; the sampling contribution to this degradation is the Moire-like phenomenon known as aliasing and the reconstruction contribution is a blurring caused by high frequency suppression. Aliasing has been investigated for line-scan systems both experimentally¹³ and theoretically¹⁴, but the effect of sample-scene phase was not addressed. Similarly this phase effect is typically ignored in studies of various reconstruction filters, e.g., reference 15. The amount and appearance of the image degradation associated with sampling and reconstruction will change as the sampling grid is shifted across the scene, i.e., as the sample-scene phase is varied. How does the magnitude of this degradation depend upon the sample-scene phase? To date a complete analysis of this question has not appeared.

In this paper the one-dimensional process of image scanning, sampling, and interpolative reconstruction is analyzed in both the spatial and frequency (Fourier) domain with the sample-scene phase appearing as an explicit, random parameter. To produce mathematically tractable results in the spatial domain it is necessary to restrict attention to simplified models of the scene, scanner PSF, and reconstruction function. Nevertheless, considerable insight into the general problem is gained with this spatial domain analysis. The results obtained in the frequency domain are completely general with no restriction on the scene frequency spectrum, scanner MTF, or reconstruction filter. This approach characterizes completely the relation of the sample-scene phase to scanning, sampling, and reconstruction and provides several measures of the resulting image degradation which can be easily evaluated by numerical quadrature. It is demonstrated that although image degradation due to sampling and reconstruction is present for all practical systems, its average magnitude can be controlled by a proper choice of scanner MTF, sampling rate, and reconstruction filter.

Although the results of this paper have general application, the primary motivation for the research was to incorporate the effect of sample-scene phase into an analysis of the image degradation due to sampling and reconstruction of remotely-sensed multispectral (e.g., Landsat) images. For such images there is known to be (see, for example, reference 16) a sample-scene phase-dependent blurring and apparent location shift of high contrast features due to sampling and reconstruction. When classifying multispectral image data, one is concerned with any blurring and shift in location of edges, lines, points, etc., since such high contrast features are commonly used as

control points in the resampling process of registering one image to another, and even small shifts in control point locations can result in significant misregistration and subsequent misclassifications¹⁷.

Formulation

Figure (1a) illustrates a one-dimensional scene, denoted $f(x-u)$, which contains a high contrast feature, in this case an edge, at the point u . A typical scanner PSF, denoted $h(x)$, is illustrated in Fig. (1b) and the image

$$g(x-u) = h(x) * f(x-u) \quad (1)$$

formed by convolving the scene with the PSF is illustrated in Fig. (1c).

The process of image sampling can be represented symbolically as the product

$$g(x-u) \text{ comb}(x) \quad (2a)$$

where sampling is accomplished by the familiar comb¹⁸ (or Shah¹⁹) function

$$\text{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x-n). \quad (2b)$$

The spatial coordinate x is normalized in units of sample interval so that image sampling occurs at the integer locations as indicated in Fig. (1d). In this coordinate system u is the sample-scene phase parameter; values of u between zero and one indicate the position (or phase) of the scene relative to the sampling grid.

Image reconstruction is commonly modeled as¹²

$$g_r(x;u) = [g(x-u) \text{ comb}(x)] * r(x) \quad (3a)$$

where the interpolation function $r(x)$ is the impulse response of the (linear shift invariant) reconstruction process. A typical interpolation function and the resultant reconstructed image $g_T(x;u)$ are illustrated in Figs. (1e) and (1f), respectively. Equation [3a] can be written equivalently as

$$g_T(x;u) = \sum_{n=-\infty}^{\infty} g(n-u)r(x-n) \quad (3b)$$

which demonstrates, consistent with the notation $g_T(x;u)$, that the reconstructed image is not simply a function of the difference $x-u$. Equation [3b] also illustrates (along with Fig. (1f)) that, for a fixed x , the extent to which neighboring image samples, $g(n-u)$, contribute to the reconstructed image is determined by the spread of the interpolation function. For digital image reconstruction (sometimes called resampling²⁰), computational considerations dictate that this spread should be small; typically $r(x)$ is identically zero for $|x|$ larger than 2 or 3.

Scanning produces an image (Fig.(1c)) which is a blurred (i.e. smoothed) copy of the original scene (Fig. (1a)); similarly, subsequent sampling and reconstruction produces additional blurring in the reconstructed image (Fig. (1f)). Thus to study the blur (i.e., the square of the radiometric error) present in a reconstructed image it is convenient to define two quantities

$$\text{image blur} = \epsilon_I^2 = \int_{-\infty}^{\infty} [\bar{f}(x-u) - g(x-u)]^2 dx \quad (4)$$

and

$$\text{sampling and reconstruction blur} = \epsilon_{SR}^2(u) = \int_{-\infty}^{\infty} [g(x-u) - g_T(x;u)]^2 dx. \quad (5)$$

These two quantities are associated with (but not equal to) the shaded areas

indicated in Figs. (2a) and (2b); ϵ_I^2 and ϵ_{SR}^2 represent the energy of the differences $f(x-u)-g(x-u)$, and $g(x-u)-g_T(x;u)$, respectively. How do ϵ_I^2 and ϵ_{SR}^2 depend upon the sample-scene phase parameter u ? For ϵ_I^2 the answer is obvious, for ϵ_{SR}^2 , it is not.

Image blur (Eq. [4]) is an inevitable but well understood effect caused solely by scanning the scene with a non-ideal (i.e., non-impulsive) PSF. Image blur is independent of the sampling and reconstruction process and changing u merely causes the scene and its image (Figs. (1a) and (1c)) to shift together along the x -axis without changing the energy of their difference. Thus image blur is independent of sample-scene phase. (This conclusion can be verified mathematically by the substitution $x' = x-u$ in Eq. [4].)

Sampling and reconstruction blur, denoted SR blur (Eq. [5]), is just as inevitable as image blur; however, unlike image blur, the amount and appearance of SR blur depends upon the sample-scene phase. Unfortunately, SR blur is not nearly as well understood as image blur, even if the sample-scene phase effect is ignored. To think of SR blur as just aliasing is to ignore the effect of imperfect reconstruction; to think of SR blur as just interpolation (or resampling) error is to ignore the effect of undersampling.

In this paper the dependence of ϵ_{SR}^2 upon sampling, reconstruction, and sample-scene phase is analyzed. In addition, ϵ_{SR}^2 is statistically characterized in terms of its mean and variance by considering an ensemble of scenes $f(x-u)$ with u as a random phase parameter uniformly distributed between zero and one. The importance of this stochastic approach is that frequently in a real scene with many high contrast features, the sample-scene phase of each is random and equally likely to be any number between zero and one. Consequently,

in a reconstructed image of this scene the SR blur associated with each high contrast feature must be analyzed as a random variable.

Spatial Analysis: Example

It is illuminating to analyze a particular image scanning, sampling, and reconstruction system in the spatial domain because the analysis is direct (although algebraically cumbersome) and the results are easily verified by digital simulation. Specifically, consider a scene $f(x-u)$ which is an edge (i.e., a step of unit height) at the location u

$$f(x-u) = \text{step}(x-u) = \begin{cases} 1 & x \geq u \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Equation [6] provides an important scene model for remote sensing applications since many images of interest are agricultural or urban and dominated by edges.

If the scanner PSF is an ideal aperture of unit response and width s , i.e.,

$$h(x) = \frac{1}{s} \text{rect}\left(\frac{x}{s}\right) = \begin{cases} \frac{1}{s} & |x| \leq \frac{s}{2} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

then the resultant blurred image of the edge is a ramp of width s and unit height, centered at u , i.e.,

$$g(x-u) = h(x) * f(x-u) = \begin{cases} 1 & x > u + \frac{s}{2} \\ \frac{x-u}{s} + \frac{1}{2} & |x-u| \leq \frac{s}{2} \\ 0 & x < u - \frac{s}{2} \end{cases} \quad (8)$$

In the coordinate system of this paper, the parameter s represents the instantaneous field of view (IFOV) of the scanning aperture in units of sample interval. From Eqs. [4], [6], and [8] the image blur in this special case is

$$\sigma_1^2 = \frac{1}{12} s \quad (9)$$

and therefore increases linearly with IFOV.

In a discussion of image blur, s should be interpreted as the scanner IFOV. However, in a discussion of SR blur it is more appropriate to interpret s as the number of samples per IFOV. Thus, for example $s = 1$ corresponds to a sampling rate at which consecutive IFOV's are just contiguous while $s = 2$ corresponds to a 50% overlap in consecutive IFOV's. Most image scanning systems operate with a sampling rate in the range $1 \leq s \leq 2$; for mathematical simplicity s is restricted to this range in all the subsequent spatial domain analysis.

Sampling the ramp image (Eq. [8]) generates image samples $g(n-u)$ which are all zero for $n \leq -1$ and all one for $n \geq 2$. The dependence of the two remaining image samples, $g(-u)$ and $g(1-u)$, upon the parameters s and u can be determined from Eq. [8]. This dependence is illustrated in Fig. (3). For u less than $1 - 0.5s$, $g(-u) = 0.5 - u/s$ and $g(1-u) = 1$; for u between $1 - 0.5s$ and $0.5s$, $g(-u) = 0.5 - u/s$ and $g(1-u) = 0.5 + (1-u)/s$; and for u greater than $0.5s$, $g(-u) = 0$ and $g(1-u) = 0.5 + (1-u)/s$.

If image reconstruction is accomplished by linear interpolation, the associated interpolation function (see Fig. (1e)) is

$$r(x) = \text{tri}(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Figure (4) illustrates the resulting reconstructed edge image as well as the original edge and its image. The shaded areas in Fig. (4b) correspond to the

SR blur that arises when $1 - 0.5s < u < 0.5s$. Figures (4a) and (4c) correspond to the other two possibilities, $0 \leq u \leq 1 - 0.5s$ and $0.5s \leq u \leq 1$, respectively. It can be shown that the dependence of SR blur upon s and u is

$$\epsilon_{SR}^2(u) = \begin{cases} \frac{1}{48s^2} [(s-2u)^2(2+2u-s)^2 + (s+2u)^2(2-2u-s)^2] & 0 \leq u \leq 1 - 0.5s \\ \frac{1}{48s^2} [(s-2u)^2(2+2u-s)^2 + (s+2u-2)^2(4-2u-s)^2] & 1 - 0.5s < u < 0.5s \\ \frac{1}{48s^2} [(s-2u)^2(2-2u+s)^2 + (s+2u-s)^2(4-2u-s)^2] & 0.5s \leq u \leq 1 \end{cases} \quad (11)$$

As Eq. [11] indicates, ϵ_{SR}^2 is non-negative for all $1 \leq s \leq 2$ and $0 \leq u \leq 1$.

Figure (5) depicts curves of ϵ_{SR}^2 versus u for various values of s .

Notice that ϵ_{SR}^2 is symmetric about $u = 0.5$ and that the shape of these curves depends dramatically upon s ; for $s = 1$ the curve is bowl-shaped with maxima at $u = 0, 1$ and a minimum at $u = 0.5$, for $s = 2$ the curve is bell-shaped with a maximum at $u = 0.5$ and minima at $u = 0, 1$. Equation [11] can be thought of as a transformation, with parameter s , of the uniformly distributed random variable u into a new random variable ϵ_{SR}^2 . The resulting probability density function (pdf) of ϵ_{SR}^2 has a very unusual shape, as illustrated in Fig. (6) for the case $s = 1$. The pdf is U-shaped indicating that the extreme values of SR blur are significantly more likely than the average value! The shape of the pdf for other values of s is similar.

Since u is uniformly distributed, the averages value of SR blur is

$$E[\epsilon_{SR}^2] = \int_0^1 \epsilon_{SR}^2 du \quad (12)$$

and the variance is

$$\text{Var}[\epsilon_{SR}^2] = \int_0^1 \left\{ \epsilon_{SR}^2 - E[\epsilon_{SR}^2] \right\}^2 du \quad (13)$$

The variance is a measure of the extent to which ϵ_{SR}^2 depends upon u and the effect of sample-scene phase dependence is negligible if and only if $\sqrt{\text{Var}[\epsilon_{SR}^2]}$ is much less than $E[\epsilon_{SR}^2]$. For example, from Fig. (5), this effect is negligible if $s = 1.5$ but not if $s = 1$.

The dependence of $E[\epsilon_{SR}^2]$ and $\text{Var}[\epsilon_{SR}^2]$ upon s can be determined algebraically from Eqs. [11], [12] and [13]. Specifically, it can be shown that

$$E[\epsilon_{SR}^2] = \frac{1}{15s^2} \quad 1 \leq s \leq 2 \quad (14)$$

a surprisingly simple result which has been verified by digital simulation. Equation [14] illustrates that for an ensemble of edges, each blurred by scanning with an ideal aperture, sampled, and reconstructed with linear interpolation, some SR blur is inevitable; this blur can be reduced by increasing the sampling rate, but it cannot be eliminated.

Unlike Eq. [14], the algebraic expression for $\text{Var}[\epsilon_{SR}^2]$ is too cumbersome to have practical value (it is an eighth degree polynomial in s divided by s^4). Instead, we present a plot of SR blur variance versus sampling rate, Fig. (7). This variance changes by three orders of magnitude over the range $1 \leq s \leq 2$ and a pronounced minimum occurs at a sampling rate of 1.5 samples per IFOV.

It should be emphasized that the previous analysis is valid only for $1 \leq s \leq 2$. Intuitively, the average SR blur will continue to decrease in the region $s > 2$ but not necessarily at the rate indicated in Eq. [14]. For small sampling rates, $s < 1$, it can be shown that

$$E[\epsilon_{SR}^2] = \frac{1}{6} - \frac{5}{18}s + \frac{1}{6}s^2 - \frac{1}{30}s^3, \quad s < 1. \quad (15)$$

Fourier Analysis

In order to include the sample-scene phase effect into the performance analysis of a general image scanning, sampling, and reconstruction system it is necessary to formulate the problem in the frequency domain. In the usual manner^{18,19}, we denote two functions $F(x)$ and $\hat{F}(v)$ as Fourier transform pairs

$$F(x) \xrightarrow{\mathcal{F}} \hat{F}(v) \quad (16a)$$

provided

$$\hat{F}(v) = \int_{-\infty}^{\infty} F(x) e^{-2\pi x v i} dx \quad (16b)$$

and

$$F(x) = \int_{-\infty}^{\infty} \hat{F}(v) e^{2\pi x v i} dv \quad (16c)$$

Since x is normalized in units of sample interval, the frequency coordinate v has units of cycles per sample interval. In this coordinate system the Nyquist (or folding) frequency is 0.5.

From Eqs. [1], [2], and [5] it follows that

$$\hat{f}(x-u) \xrightarrow{\mathcal{F}} e^{-2\pi u v i} \hat{f}(v) \quad (17a)$$

$$g(x-u) \xrightarrow{\mathcal{F}} e^{-2\pi u v i} \hat{h}(v) \hat{f}(v) \quad (17b)$$

and

$$g_r(x;u) \xrightarrow{\mathcal{F}} \hat{r}(v) \sum_{n=-\infty}^{\infty} e^{-2\pi u (v-n) i} \hat{h}(v-n) \hat{f}(v-n) \quad (17c)$$

where $\hat{f}(v)$ is the (zero phase) scene frequency spectrum, $\hat{h}(v)$ is the scanner OTF, $\hat{r}(v)$ is the reconstruction filter (i.e., the Fourier transform of the interpolation function), and as before u is the (random) sample-scene phase

parameter. Equations [17a] and [17b] are immediate consequences of the shift and convolution theorems. To derive Eq. [17c], the convolution theorem applied to Eq. [3a] yields

$$g_r(x;u) \xrightarrow{\mathcal{F}} \left\{ \left[e^{-2\pi i u v} \hat{h}(v) \hat{f}(v) \right] * \text{comb}(v) \right\} \hat{r}(v), \quad (18)$$

a result which uses Eq. [17b] and the fact that the comb function has the curious property of being its own Fourier transform¹⁸. Convolution with the comb function in Eq. [18] produces the image spectrum replication in Eq. [17c]; this replication is the inevitable result of sampling.

In the previous spatial domain analysis the scanner PSF (Eq. [7]) and interpolation function (Eq. [10]) were normalized to unit area. That normalization is used throughout this paper; in the frequency domain it means that at zero frequency

$$\hat{h}(0) = \int_{-\infty}^{\infty} h(x) dx = 1 \quad (19a)$$

and

$$\hat{r}(0) = \int_{-\infty}^{\infty} r(x) dx = 1. \quad (19b)$$

Parseval's theorem applied to Eqs. [4], [17a], and [17b] yields the image blur, i.e.,

$$\epsilon_I^2 = \int_{-\infty}^{\infty} |1 - \hat{h}(v)|^2 |\hat{f}(v)|^2 dv \quad (20)$$

which illustrates (again) that image blur is independent of sample-scene phase. If the scanner PSF is an impulse, then $\hat{h}(v) = 1$ for all v and image blur is zero for all scenes. For all practical PSF's the condition $\hat{h}(v) = 1$

can not hold for all v ; at high frequencies $\hat{h}(v)$ approaches zero. As Eq. [20] indicates, the cause of image blur is the presence of significant scene energy, $|\hat{f}(v)|^2$, at frequencies where $\hat{h}(v) \neq 1$.

An expression for SR blur analogous to Eq. [20] can be obtained from Eqs. [5], [17b] and [17c]; however a somewhat different approach is preferable. This approach is based upon the observation that ϵ_{SR}^2 is periodic in u with period one, and thus

$$\epsilon_{SR}^2(u) = \sum_{m=-\infty}^{\infty} a_m e^{2\pi u m i} \quad (21a)$$

where the Fourier coefficients are given by

$$a_m = \int_0^1 \epsilon_{SR}^2 e^{-2\pi u m i} du. \quad (21b)$$

The coefficient a_0 has particular significance; it is the average SR blur, i.e., from Eq. [12]

$$E[\epsilon_{SR}^2] = a_0. \quad (22)$$

In addition, it follows from Eqs. [13] and [21a] that SR blur variance is

$$\text{Var}[\epsilon_{SR}^2] = 2 \sum_{m=1}^{\infty} |a_m|^2. \quad (23)$$

In general for $m \neq 0$ the Fourier coefficients are complex. However because ϵ_{SR}^2 is real, they satisfy $a_{-m} = a_m^*$ for $m = 1, 2, 3, \dots$ where $(\cdot)^*$ denotes complex conjugate.

The explicit dependence of the Fourier coefficients (and thus $E[\epsilon_{SR}^2]$ and $\text{Var}[\epsilon_{SR}^2]$) upon the scene spectrum, scanner OTF, and reconstruction filter can

be determined by first recognizing that, from Eqs. [17b] and [17c]

$$g(x-u) - g_r(x;u) \xrightarrow{\mathcal{F}} e^{-2\pi u v i} \sum_{n=-\infty}^{\infty} (\delta_n - \hat{r}(v)) \hat{h}(v-n) \hat{f}(v-n) e^{2\pi u n i} \quad (24)$$

where in the summation $\delta_n = 1$ if $n = 0$ and $\delta_n = 0$ otherwise. A combination of Parseval's equation, Eqs. [5], [21b] and [24] yields an imposing looking expression for a_m

$$a_m = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \sum_{\bar{n}=-\infty}^{\infty} (\delta_n - \hat{r}(v)) (\delta_{\bar{n}} - \hat{r}^*(v)) \hat{h}(v-n) \hat{h}^*(v-\bar{n}) \hat{f}(v-n) \hat{f}^*(v-\bar{n}) \\ \times \left[\int_0^1 e^{-2\pi u(m-n+\bar{n})i} du \right] dv \quad (25)$$

Fortunately, because of the integration with respect to u , the only non-zero terms in the double summation are those for which $n-\bar{n} = m$. With some mathematical manipulation, which is omitted for brevity, Eq. [25] can be reduced to its simplest form

$$a_m = \int_{-\infty}^{\infty} \left[\delta_m - \hat{r}(v) - \hat{r}^*(v-m) + \sum_{n=-\infty}^{\infty} |\hat{r}(v-n)|^2 \right] \hat{h}^*(v) \hat{h}(v-m) \hat{f}^*(v) \hat{f}(v-m) dv \quad (26)$$

In summary, for a general image scanning, sampling and reconstruction system the image blur is given by Eq. [20] and the SR blur is given by a Fourier series (Eq. [21a]) in the random phase parameter u with coefficients given by Eq. [26]. The mean and variance of SR blur are given by Eqs. [22] and [23]. All of these quantities can be calculated by numerical quadrature provided values of the scene frequency spectrum $\hat{f}(v)$, the scanner OTF $\hat{h}(v)$, and the interpolation filter $\hat{r}(v)$ are available.

As an example of this Fourier analysis, if the scene is an edge (Eq. [6]), the scanner PSF is an ideal aperture (Eq. [7]), and the interpolation function

is linear (Eq. [10]), then

$$\hat{f}(v) = \frac{1}{2} \delta(v) + \frac{1}{i2\pi v} \quad (27a)$$

$$\hat{h}(v) = \text{sinc}(sv) \quad (27b)$$

and

$$\hat{r}(v) = \text{sinc}^2(v). \quad (27c)$$

It can be shown that the associated image blur is

$$\varepsilon_I^2 = \int_{-\infty}^{\infty} \frac{[1 - \text{sinc}(sv)]^2}{4\pi^2 v^2} dv \quad (28)$$

and the SR blur is

$$\varepsilon_{SR}^2(u) = E[\varepsilon_{SR}^2] + 2 \sum_{m=1}^{\infty} a_m \cos(2\pi mu) \quad (29a)$$

where

$$E[\varepsilon_{SR}^2] = \int_{-\infty}^{\infty} \left[1 - 2\text{sinc}^2(v) + \sum_{n=-\infty}^{\infty} \text{sinc}^4(v-n) \right] \frac{\text{sinc}^2(sv)}{4\pi^2 v^2} dv \quad (29b)$$

and where the Fourier coefficients are real and given by .

$$a_m = \int_{-\infty}^{\infty} \left[-\text{sinc}^2(v) - \text{sinc}^2(v-m) + \sum_{n=-\infty}^{\infty} \text{sinc}^4(v-n) \right] \times \frac{\text{sinc}(sv)\text{sinc}[s(v-m)]}{4\pi^2 v(v-m)} dv \quad m = 1, 2, \dots \quad (29c)$$

Because of the singularity in the edge spectrum (Eq. [27a]) at $v = 0$, some caution must be exercised in the numerical integration of Eqs. [28], [29b], and [29c]. However, the singularities in the integrands are removable and present no fundamental difficulties.

SR Blur

Considerable insight into SR blur can be obtained without resorting to numerical simulation. In particular, from Eqs. [22] and [26] the average SR blur is given by

$$E[\epsilon_{SR}^2] = \int_{-\infty}^{\infty} e^2(\nu) |\hat{h}(\nu) \hat{f}(\nu)|^2 d\nu \quad (30)$$

where $e^2(\nu)$ is the non-negative function defined by

$$e^2(\nu) = 1 - 2\Re[\hat{r}(\nu)] + \sum_{n=-\infty}^{\infty} |\hat{r}(\nu-n)|^2 \quad (31a)$$

$$= |1 - \hat{r}(\nu)|^2 + \sum_{n \neq 0} |\hat{r}(\nu-n)|^2 \quad (31b)$$

and $\Re[\cdot]$ denotes the real part of a complex variable. Since ϵ_{SR}^2 is a non-negative random variable, it is identically zero if and only if its mean is zero. From Eq. [30], $E[\epsilon_{SR}^2] = 0$ if and only if the integrand is identically zero for all ν . In other words, for a general image scanning, sampling and reconstruction system some SR blur is inevitable unless

$$e(\nu) |\hat{h}(\nu) \hat{f}(\nu)| = 0 \quad (32)$$

for all ν .

For a well designed scanning system it is common to use electronic filtering prior to sampling and thereby shape the scanner MTF, $|\hat{h}(\nu)|$, so that the image spectrum is band-limited, i.e.,

$$|\hat{h}(\nu) \hat{f}(\nu)| = 0 \quad (33)$$

for all $|\nu| > \nu_c$. For such a system, Eq. [32] will be satisfied at high frequencies. However, below the cut-off frequency, ν_c , Eq. [32] can be satisfied only by choosing the reconstruction filter $\hat{r}(\nu)$ so that $e^2(\nu)$ is zero.

What choice of reconstruction filter will satisfy the equation

$$e^2(v) = 0, \quad |v| \leq v_c \quad (34)$$

for all $|v| \leq v_c$? Because of the first term in Eq. [31b], $\hat{r}(v)$ must be one for low frequencies; because of the remaining (side-band) terms $\hat{r}(v)$ must be zero for high frequencies. In fact, in terms of v_c , the solution of Eq. [34] is completely characterized as follows. If $v_c > 0.5$, Eq. [34] has no solution, and if $v_c \leq 0.5$, the only solution to Eq. [34] is

$$\hat{r}(v) = \begin{cases} 1 & |v| \leq v_c \\ \text{arbitrary} & v_c < |v| \leq 0.5 \\ 0 & |v| > 0.5 \end{cases} \quad (35)$$

This discussion establishes what amounts to a sampling theorem²¹ for a general image scanning, sampling, and reconstruction system. Namely, there is no SR blur (Eq. [5]) if and only if all the following are true:

- (i) the image spectrum is band-limited;
- (ii) the cut-off frequency v_c is 0.5 cycles per sample interval or less;
- (iii) the reconstruction filter satisfies Eq. [35].

Condition (ii) represents the Nyquist criteria for sufficient sampling, i.e., it determines an upper bound for the sample interval size. Conditions (i) and (iii) are idealizations; a truly band-limited image spectrum is convenient to assume but difficult to achieve and all finite width reconstruction filters are "non-ideal," i.e., they do not satisfy Eq. [35]. Thus, for all practical systems some SR blur is inevitable.

There are three effects which contribute to SR blur: imperfect reconstruction, aliasing, and sample-scene phase dependence. To see the contribution of each, we can combine Eqs. [21a] and [22] to yield

$$\varepsilon_{SR}^2(u) = E[\varepsilon_{SR}^2] + \sum_{m \neq 0} a_m e^{2\pi i u m} \quad (36)$$

From Eqs. [30] and [31b], $E[\varepsilon_{SR}^2]$ can be written as the sum of two terms, i.e.,

$$E[\varepsilon_{SR}^2] = \varepsilon_R^2 + \varepsilon_S^2 \quad (37a)$$

where

$$\varepsilon_R^2 = \int_{-\infty}^{\infty} |1 - \hat{r}(v)|^2 |\hat{h}(v)\hat{f}(v)|^2 dv \quad (37b)$$

and

$$\varepsilon_S^2 = \int_{-\infty}^{\infty} \left[\sum_{n \neq 0} |\hat{r}(v-n)|^2 \right] |\hat{h}(v)\hat{f}(v)|^2 dv. \quad (37c)$$

The term ε_R^2 accounts for imperfect reconstruction; it is analogous to image blur, Eg. [20], and it measures the contribution to SR blur caused by the presence of significant image energy, $|\hat{h}(v)\hat{f}(v)|^2$, at frequencies where $\hat{r}(v) \neq 1$. The term ε_S^2 accounts for aliasing due to undersampling; it measures the contribution to SR blur caused by the presence of significant image energy at frequencies where the energy in the reconstruction filter side-bands, $\sum_{n \neq 0} |\hat{r}(v-n)|^2$, is not zero. An equivalent, but more familiar expression¹⁴ for ε_S^2 follows from the identity

$$\varepsilon_S^2 = \int_{-\infty}^{\infty} \left[\sum_{n \neq 0} |\hat{r}(v-n)|^2 \right] |\hat{h}(v)\hat{f}(v)|^2 dv = \int_{-\infty}^{\infty} |\hat{r}(v)|^2 \left[\sum_{n \neq 0} |\hat{h}(v-n)\hat{f}(v-n)|^2 \right] dv. \quad (37d)$$

Equation [37d] demonstrates the duality between side-band energy in the reconstruction filter and image spectrum. Both ε_R^2 and ε_S^2 are independent of u and correctly account for the sample-scene phase-averaged effects of imperfect reconstruction and aliasing.

From Eqs. [36] and [37a], ε_{SR}^2 can be written

$$\varepsilon_{SR}^2(u) = \varepsilon_S^2 + \varepsilon_R^2 + \psi(u) \quad (38a)$$

where

$$\phi(u) = \sum_{m \neq 0} a_m e^{2\pi u m i} = 2 \sum_{m=1}^{\infty} \Re [a_m e^{2\pi u m i}] \quad (38b)$$

The term ϕ is the sample-scene phase dependent contribution to SR blur; it is a real valued random variable with zero mean and variance $2 \sum_{m=1}^{\infty} |a_m|^2$. Phase dependence is zero if and only if $a_m = 0$ for all $m = 1, 2, \dots$. For a particular system this phase dependence is best investigated by numerical simulation using Eqs. [26] and [38b]. However, one condition sufficient to guarantee that $\phi = 0$ can be obtained by inspection; the integrand in Eq. [26] will be zero for all v if the image spectrum is band-limited with the cut-off frequency less than 0.5. That is, sufficient sampling of a (truly) band-limited image will eliminate the sample-scene phase dependence in SR blur. Of course, some SR blur will still be present unless the reconstruction filter satisfies Eq. [35].

An inspection of Eq. [27] reveals why the example considered previously in the spatial domain analysis exhibited a non-zero, phase dependent, SR blur for all sampling rates, s ; the image spectrum is not band-limited and the reconstruction filter is imperfect. For this example both

$$\epsilon_R^2 = \int_{-\infty}^{\infty} [1 - \text{sinc}^2(v)]^2 \left[\frac{\text{sinc}(sv)}{2\pi v} \right]^2 dv \quad (39a)$$

and

$$\epsilon_S^2 = \int_{-\infty}^{\infty} \left[\sum_{n \neq 0} \text{sinc}^4(v-n) \right] \left[\frac{\text{sinc}(sv)}{2\pi v} \right]^2 dv \quad (39b)$$

are positive for all s . In addition, it can be verified numerically (using Eqs. [23] and [29c]) that $\text{Var}[\epsilon_{SR}^2] > 0$ and thus the phase dependence, $\phi(u)$, becomes small, but does not disappear as s becomes large. See also Fig. (7) which illustrates the magnitude of this phase dependence for $1 \leq s \leq 2$.

Reconstruction Filters

To summarize some of the previous discussion, Eq. [31b] defines $e^2(v)$ in terms of the reconstruction filter $\hat{r}(v)$. As Eq. [30] reveals, the cause of SR blur is the presence of significant image energy at frequencies where $e^2(v)$ is not zero. Just as the scanner OTF can be interpreted as a pre-sampling filter acting upon the scene Wiener spectrum, $|\hat{f}(v)|^2$, to produce image blur (Eq. [20]), the reconstruction filter can be interpreted as a post-sampling filter acting upon the image Wiener spectrum, $|\hat{h}(v)\hat{f}(v)|^2$, to produce average SR blur (Eq. [30]). In other words, $|1-\hat{h}(v)|^2$ is to image blur as $e^2(v)$ is to average SR blur. It should be emphasized that $e^2(v)$ is not just $|1-\hat{r}(v)|^2$; sampling causes the appearance of the side band terms $\sum_{n \neq 0} |\hat{r}(v-n)|^2$ in Eq. [31b]. If $r(x)$ is the ideal interpolation function $\text{sinc}(x)$, then $\hat{r}(v) = \text{rect}(v)$ and $e^2(v) = 2 - 2\text{rect}(v)$. What does $e^2(v)$ look like for some common (non-ideal) reconstruction filters?

Table I lists the characteristics of three common digital interpolation functions, nearest neighbor, linear, and cubic and the ideal interpolation function, sinc. There are actually several variants of the cubic interpolator in the literature, each derived as a finite-width approximation to $\text{sinc}(x)$; the form we have chosen is due to Bernstein²². Figure (3) is a plot of $e^2(v)$ versus v for the four interpolation functions. A comparison of the relative magnitude of the four curves in the region $0 \leq v \leq 0.5$ yields the conclusion that for any band-limited and sufficiently sampled image: (i) the ideal interpolator causes no SR blur; (ii) of the three common digital interpolation functions, cubic causes the least SR blur and; (iii) nearest neighbor interpolation causes the most SR blur. (Unfortunately, it is necessary to qualify conclusion (ii) because of the small hump on the cubic curve at low

frequency. It would be mathematically possible to construct a band-limited and sufficiently sampled image, with virtually all its energy at $\nu = 0.05$, for which linear interpolation causes slightly less SR blur than cubic.) Since the $e^2(\nu)$ curves cross above the Nyquist frequency (0.5), if the image is not band-limited and sufficiently sampled, conclusions (i), (ii), and (iii) are not necessarily valid. In fact, in this case conclusion (i) is invalid and the ideal interpolator may cause more average SR blur than some of the others!

The example we have considered throughout this paper--edge reconstruction--has an image spectrum which is not band-limited. For this example the average SR blur is

$$E[e_{SR}^2] = \int_{-\infty}^{\infty} \frac{e^2(\nu) \text{sinc}^2(s\nu)}{4\pi^2\nu^2} d\nu \quad (40)$$

where $e^2(\nu)$ is given by Eq. [31] and the terms $\hat{r}(\nu)$ and $\sum_{n=-\infty}^{\infty} |\hat{r}(\nu-n)|^2$ are listed in Table I for each interpolation function. Equation [40] has been evaluated numerically for various values of the sampling rate, s , and the results are illustrated in Fig. (9). Of the three common digital interpolators, nearest neighbor is clearly the worst choice (as expected) and cubic is the best choice except at low sampling rates, $s \lesssim 0.6$. In this example, when s is less than 0.3, the image spectrum has so much energy above the Nyquist frequency that the ideal interpolator actually causes more average SR blur than either linear or cubic. In fact, cubic causes less average SR blur than sinc for values of s as large as 1.1.

Table I lists, for each interpolator, a closed form expression for the infinite series $\sum_{n=-\infty}^{\infty} |\hat{r}(\nu-n)|^2$. This expression can be used to simplify considerably the numerical evaluation of $e^2(\nu)$ and is based upon the identity

$$\sum_{n=-\infty}^{\infty} |\hat{r}(v-n)|^2 = (r^*r)(0) + 2 \sum_{n=1}^{\infty} (r^*r)(n) \cos(2\pi nv) \quad (41)$$

Equation [41] is valid for any interpolation function provided it is even, i.e., $r(x) = r(-x)$ for all x . For all practical interpolation functions, $(r^*r)(n)$ will be zero except at small n and the infinite series in Eq. [41] can be evaluated in closed form. The validity of Eq. [41] follows from the observations that if $r(x)$ is even, $\hat{r}(v)$ will be real, $r(x)^*r(x)$ will be even, and

$$\sum_{n=-\infty}^{\infty} |\hat{r}(v-n)|^2 = \hat{r}^2(v) * \text{comb}(v) = \sum_{n=-\infty}^{\infty} (r^*r)(n) e^{-2\pi n v i}. \quad (42)$$

The last equality in Eq. [42] can be verified by taking the Fourier transform of each side. In Eqs. [41] and [42], $(r^*r)(n)$ represents $r(x)^*r(x)$ evaluated at $x = n$.

The assumption that the interpolation function $r(x)$ is even is not restrictive; it is difficult to imagine any situation in which $r(x)$ would be constructed otherwise. It can be shown that if $r(x)$ is even, then $\hat{r}(v)$, $\sum_n |\hat{r}(v-n)|^2$, and $e^2(v)$ will all be even functions of v .

Discussion

The emphasis throughout this paper has been on an analysis of the degradation, $e_{SR}^2(u)$, caused by image sampling and reconstruction. The analysis is unique in that it includes the sample-scene phase as an explicit, random parameter. Results established in the frequency domain provide a complete characterization of $e_{SR}^2(u)$ as the sum of two terms, $E[e_{SR}^2]$ and $\phi(u)$. The first term accounts for the mean effect of sampling and reconstruction averaged over all

sample-scene phases, the second term accounts for variations about this mean.

If the image is band-limited and sufficiently sampled, the variations about the mean are suppressed, i.e., $c(u) = 0$. If, in addition, the reconstruction filter is ideal, the mean is zero and SR blur disappears.

Table II is a logic flow diagram which summarizes the results of the frequency domain analysis. To avoid unnecessary clutter in this diagram, only a band-limited image is considered and the notation

$$\hat{g}(v) = \hat{h}(v)\hat{f}(v) \quad (43a)$$

$$v(v;m) = \sum_{n=-\infty}^{\infty} |\hat{r}(v-n)|^2 - \hat{r}(v) - \hat{r}^*(v-m) \quad (43b)$$

is introduced to simplify some equations. Since all spatial and frequency variables in this paper are referenced to a coordinate system in which the sample interval has unit length, all frequencies (v) in Table II have units of cycles per sample interval and the Nyquist frequency is 0.5.

To better understand the results of this paper, it is important to appreciate the inverse relation between the parameters v_c and s . For a general scanner PSF (not just an ideal aperture, Eq. [7]), s represents the width of the IFOV relative to the sample interval or, equivalently, the number of samples per IFOV. Just as s is proportional to the effective width of the scanner PSF, the cut-off frequency, v_c , is proportional to the effective width of the scanner CTF. Because of the scaling property of Fourier transform pairs, it follows that

$$s v_c = \text{constant}. \quad (44)$$

Thus, for example, reducing the cut-off frequency by 50% is equivalent to doubling the sampling rate.

From Eq. [30] it can be seen that, for a fixed sample interval size, there are two ways to reduce SR blur. One way is to shape the reconstruction filter so that $\epsilon^2(v)$ is small at those frequencies where $|\hat{h}(v)\hat{f}(v)|^2$ is large. As Fig. (8) indicates, this technique of post-sample filtering can be quite effective, but only if the IFOV width, s , is large enough so that $v_c < 0.5$. Reducing SR blur by shaping the reconstruction filter does not effect image blur. The other way to reduce SR blur is to shape the scanner MTF so that $|\hat{h}(v)|^2$ is small at those frequencies where $\epsilon^2(v)|\hat{f}(v)|^2$ is large. This technique of pre-sample filtering definitely does effect image blur. In fact, there is a trade-off: reducing SR blur by shaping the scanner MTF inevitably increases image blur.

Figure (9) illustrates the trade-off between image blur and SR blur in the special case of an edge scanned with an ideal aperture. For a fixed sample interval size, as s increases the IFOV is broadened, the effective cut-off frequency is reduced, and high frequencies are suppressed. The result is an increase in image blur and a decrease in SR blur.

Finally, it is important to note that the total blur (i.e., the square of the total radiometric error) associated with the entire process of scanning, sampling and reconstruction is

$$\epsilon^2(u) = \int_{-\infty}^{\infty} [f(x-u) - g_r(x;u)]^2 dx. \quad (45)$$

This sample-scene phase dependent blur is not just the sum of ϵ_I^2 and $\epsilon_{SR}^2(u)$. However, from the triangle inequality, it follows that

$$\epsilon^2(u) \leq \epsilon_I^2 + \epsilon_{SR}^2(u) \quad (46a)$$

and

$$E[\epsilon^2] \leq \epsilon_I^2 + E[\epsilon_{SR}^2]. \quad (46b)$$

Therefore, the results of this paper can be used directly to establish an upper bound on ε^2 and $E[\varepsilon^2]$.

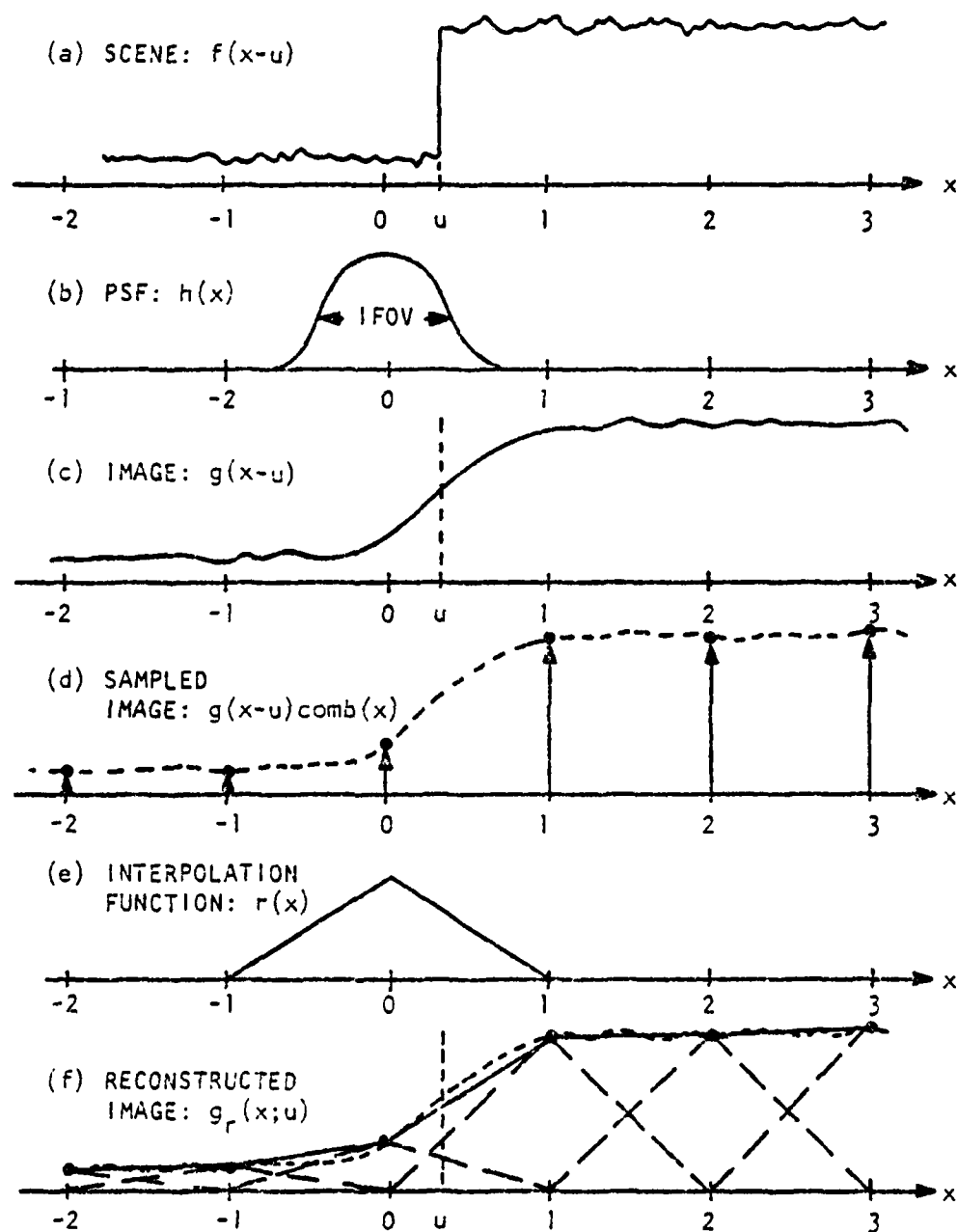


Fig. 1.

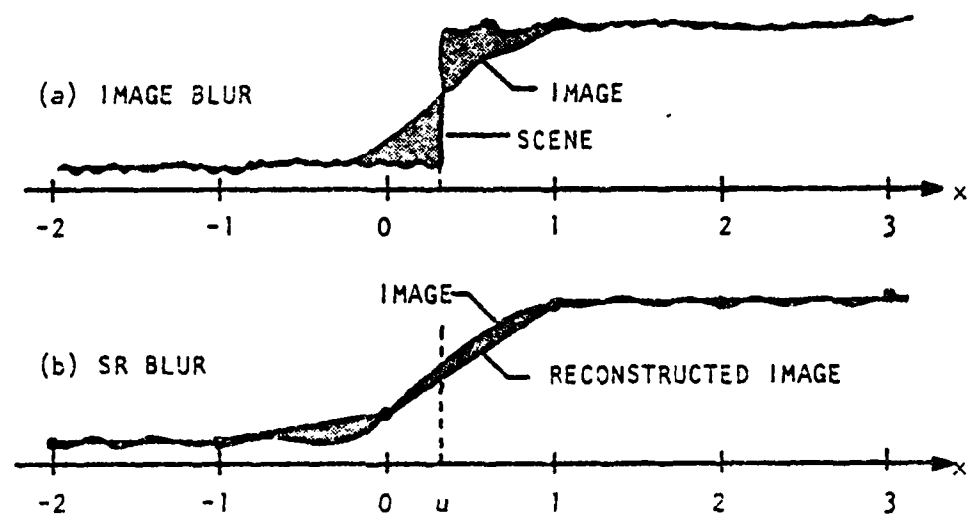


Fig. 2.

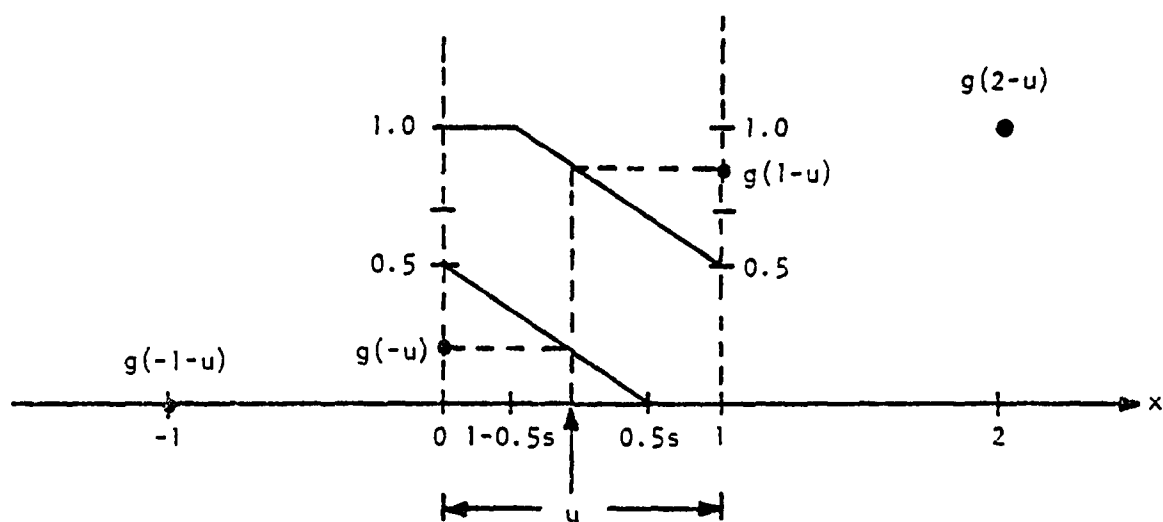


Fig. 3.

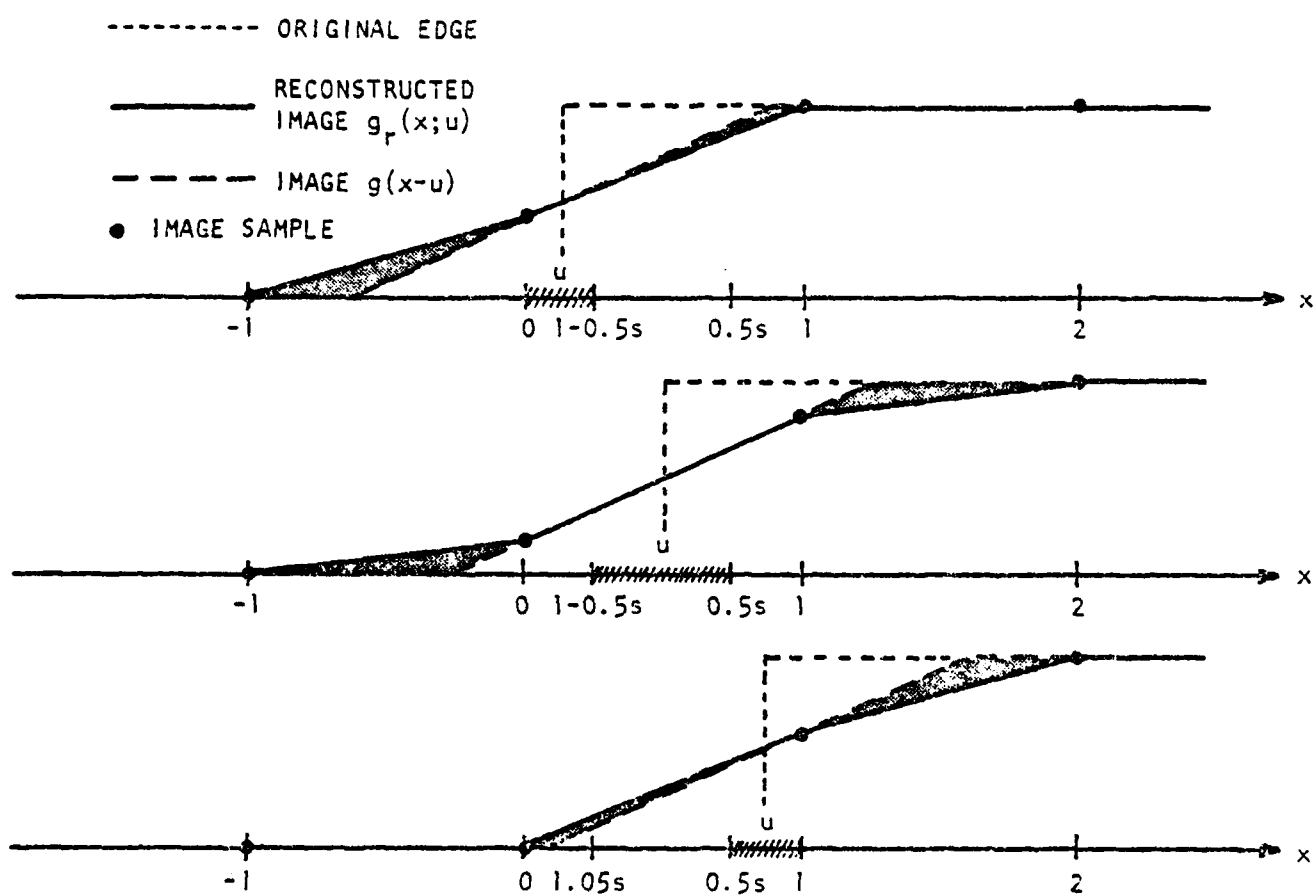


Fig. 4.

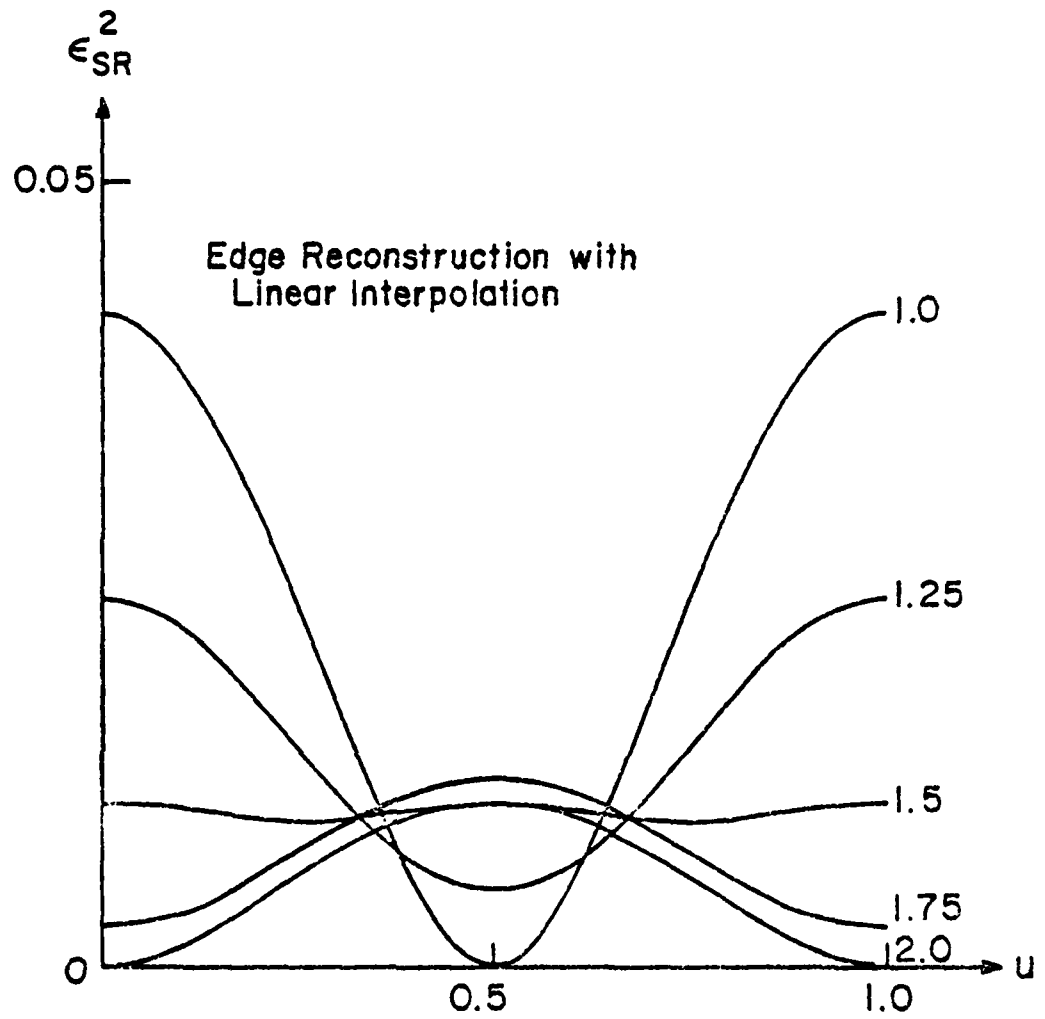


Fig. 5.

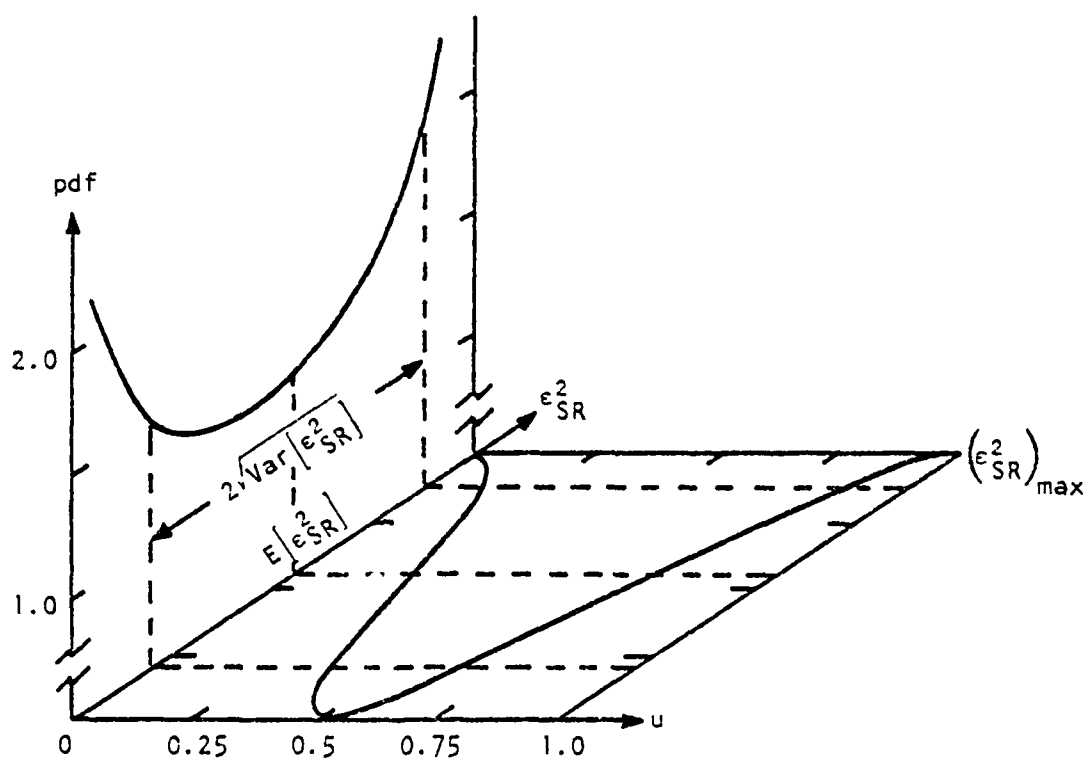


Fig. 6.

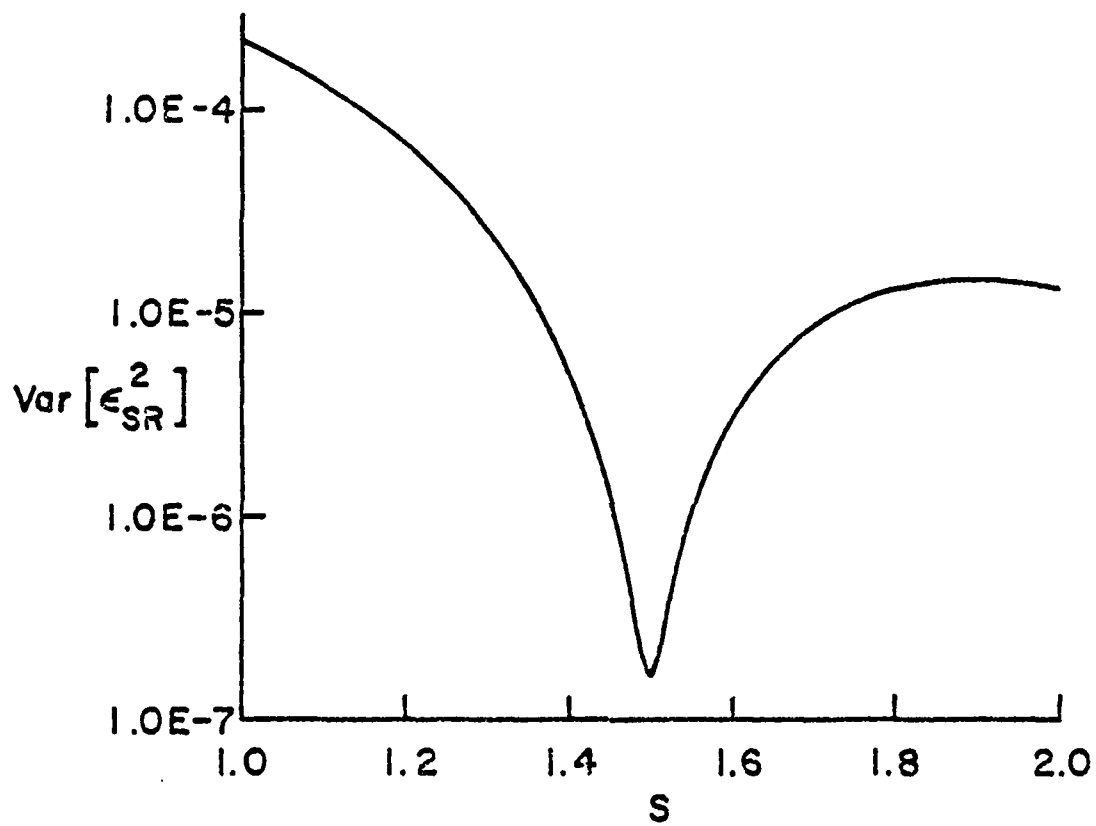


Fig. 7.

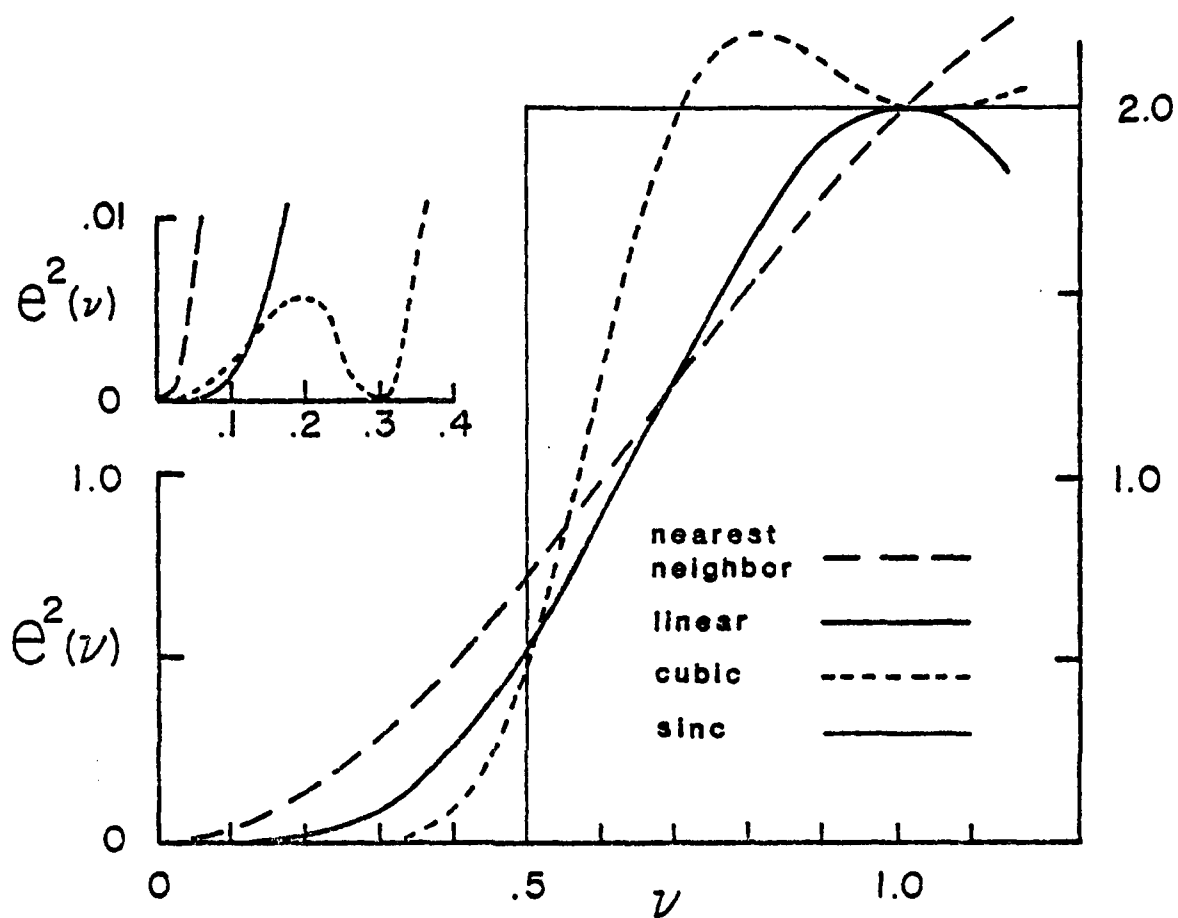


Fig. 8.

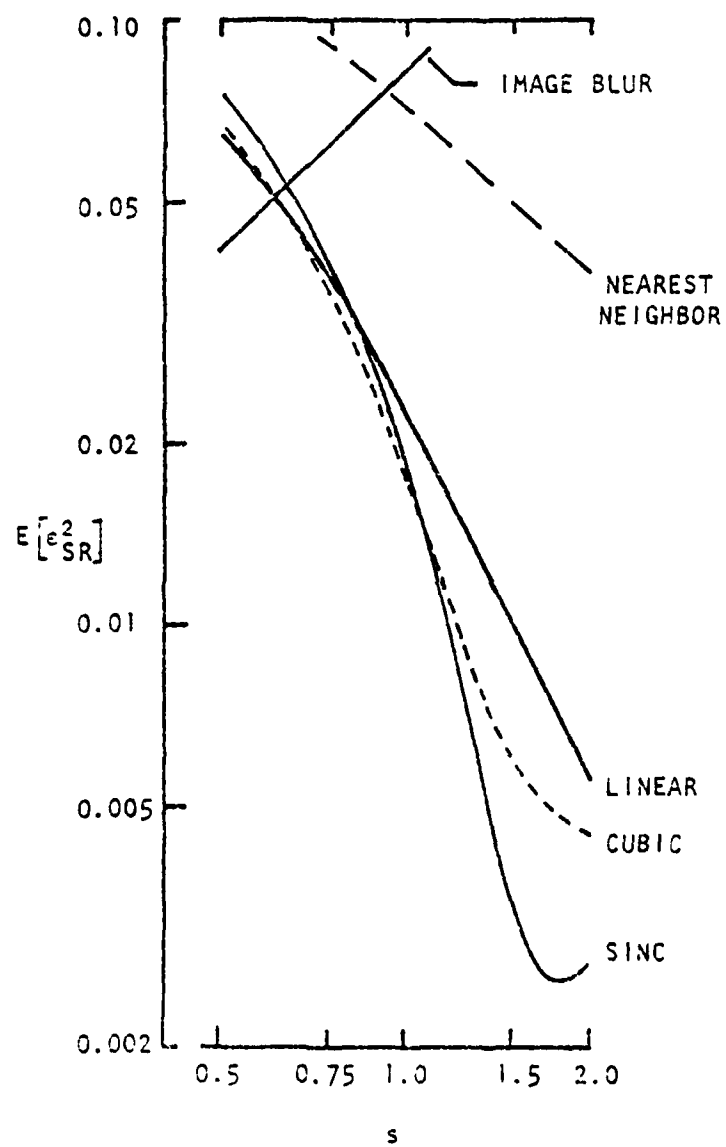
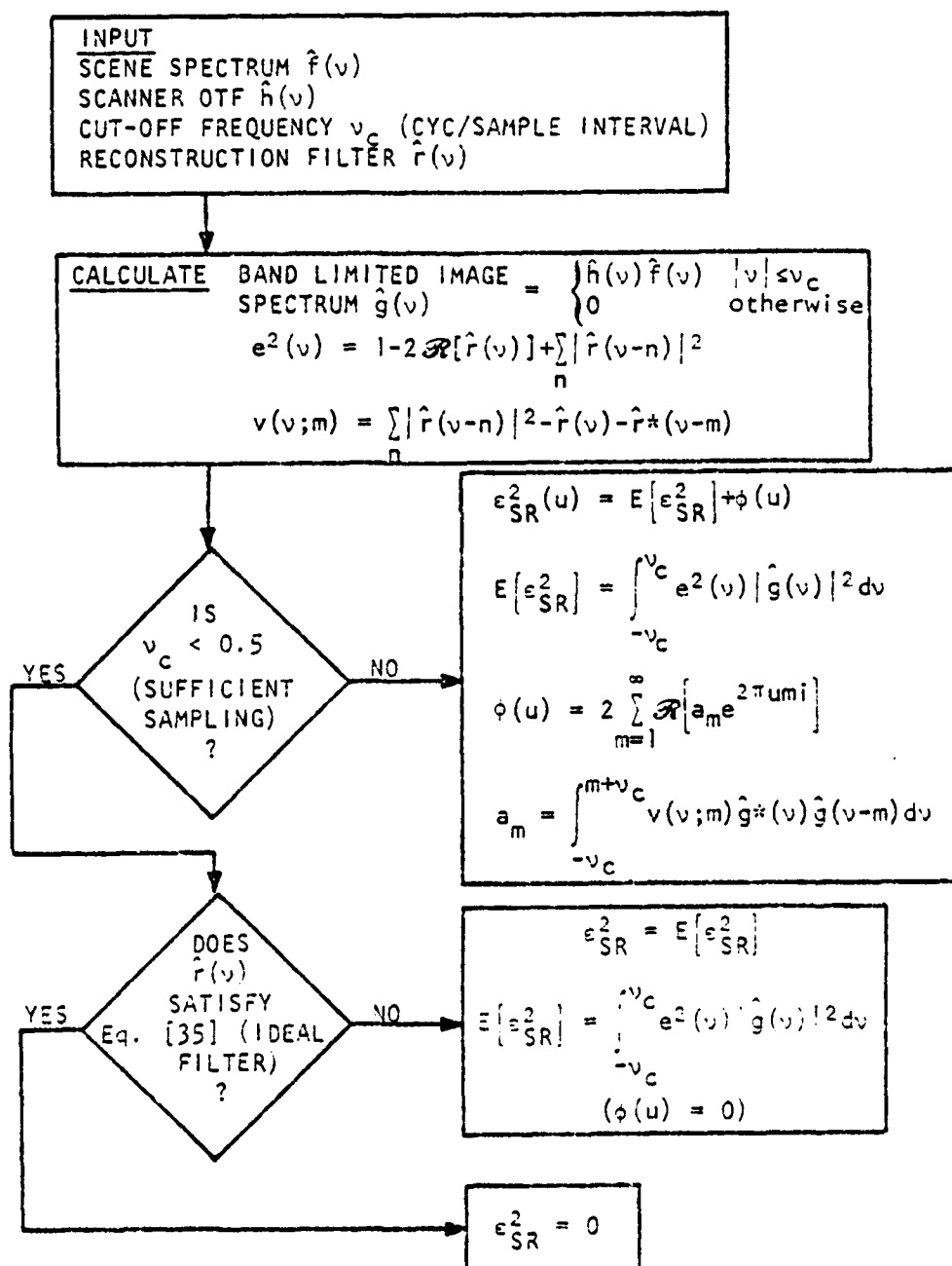


Fig. 9.

Table 1. Characteristics of three common digital reconstruction functions, nearest neighbor, linear, cubic, and the ideal interpolator, sinc.

INTERPOLATION FUNCTION, $r(x) \xrightarrow{\mathcal{F}} \text{RECONSTRUCTION FILTER, } \hat{r}(v)$				
	NEAREST NEIGHBOR	LINEAR	CUBIC	SINC
$r(x)$	1 $ x < 0.5$ 0 otherwise ($\text{rect}(x)$)	1 - $ x $ $ x < 1$ 0 otherwise ($\text{tri}(x)$)	1 - 2 $ x ^2$ + $ x ^3$ $ x < 1$ (1 - $ x $)(2 - $ x $) ² $1 \leq x \leq 2$ 0 otherwise	$\frac{\sin(\pi x)}{\pi x}$ ($\text{sinc}(x)$)
$\hat{r}(v)$	$\text{sinc}(v)$	$\text{sinc}^2(v)$	$\frac{\text{sinc}(2v) + 2\text{sinc}(4v) - 6 \left[\frac{1 - \cos(2\pi v)}{(2\pi v)^2} \right] \cos(2\pi v)}{(\pi v)^2}$	$\text{rect}(v)$
$\sum_{n=-\infty}^{\infty} \hat{r}(v-n) ^2$	1	$\frac{2 + \cos(2\pi v)}{3}$	$\frac{190 + 51 \cos(2\pi v) - 34 \cos(4\pi v) + 3 \cos(6\pi v)}{210}$	1

Table II. The calculation of SR blur. The image spectrum, $\hat{g}(v)$, is assumed to be zero for all $|v| > v_c$.



LIST OF FIGURE CAPTIONS

- Fig. 1. The image scanning, sampling, and reconstruction process.
- Fig. 2. (a) The blurring introduced by scanning. (b) The additional blurring introduced by sampling and reconstruction.
- Fig. 3. Ramp image sample dependence upon the sample-scene phase parameter.
- Fig. 4. SR blur dependence upon the sample-scene phase parameter. Three distinct situations are depicted corresponding to (a) $0 \leq u < 1 - 0.5s$, (b) $1 - 0.5s \leq u \leq 0.5s$, and (c) $0.5 < u \leq 1$, respectively.
- Fig. 5. The dependence of SR blur upon the sample-scene phase parameter. The curves correspond to $s=1.0, 1.25, 1.50, 1.75$, and 2.00 , respectively. For $s=1.50$, ϵ_{SR}^2 is nearly constant.
- Fig. 6. The probability density (vertical axis) of ϵ_{SR}^2 for the case $s=1$. The extreme values of ϵ_{SR}^2 are significantly more likely than the average value, $E[\epsilon_{SR}^2]$.
- Fig. 7. SR blur variance versus sampling rate (samples/IFOV) for edge reconstruction.
- Fig. 8. The function $e^2(v)$ for three common interpolators, nearest neighbor, linear, cubic and the ideal interpolator, sinc. The insert illustrates the small hump at low frequencies for cubic.
- Fig. 9. Average SR blur as a function of sampling rate (samples/IFOV) for the interpolators, nearest neighbor, linear, cubic and sinc. The image is an edge scanned with an ideal aperture.

LIST OF TABLES

Table I. Characteristics of three common digital reconstruction functions, nearest neighbor, linear, cubic, and the ideal interpolator, sinc.

Table II. The calculation of SR blur. The image spectrum, $\hat{g}(v)$, is assumed to be zero for all $|v| > v_c$.

REFERENCES

1. P. Mertz and F. Gray, Bell Syst. Tech. J. 13, 464 (1934).
2. O. H. Schade, J. SMPTE 56, 137 (1951); J. SMPTE 58, 181 (1952); J. SMPTE 61, 97 (1953); J. SMPTE 64, 593 (1955).
3. M. W. Baldwin, Jr., Bell Syst. Tech. J. XIX, 563 (1940).
4. R. D. Kell, A. V. Bedford, and G. L. Fredendall, RCA Rev. V, 8 (1940).
5. O. H. Schade, Sr., J. SMPTE 73, 81 (1964).
6. L. G. Callahan and W. M. Brown, Appl. Opt. 2, 401 (1963).
7. A. Macovski, Appl. Opt. 9, 1906 (1970).
8. A. H. Robinson, Appl. Opt. 12, 2344 (1973).
9. F. O. Huck and S. K. Park, Appl. Opt. 14, 2508 (1975).
10. R. A. Gonsalves and P. S. Considine, Opt. Eng. 15, 64 (1976).
11. L. M. Biberman, ed., *Perception of Displayed Information*, (Plenum Press, New York, 1973).
12. W. K. Pratt, *Digital Image Processing*, (John Wiley & Sons, New York, 1978).
13. S. J. Katzberg, F. O. Huck and S. D. Wall, Appl. Opt. 12, 1054 (1973).
14. F. O. Huck, S. K. Park, and N. Halyo, Appl. Opt. 19, 2174 (1980).
15. R. J. Arguello, Proc. SPIE 271, 86 (1981).
16. W. W. Marshall and E. Behane, Mead Tech. Lab. Final Report, 75-3 (1975).
17. F. C. Billingsley, Jet Propulsion Lab. Publ., p. 81 (1981).

18. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, (John Wiley & Sons, New York, 1978).
19. R. Bracewell, *The Fourier Transform and Its Applications*, (McGraw-Hill Book Co., New York, 1965).
20. S. Shlien, *Can. J. Rem. Sensing* 5, 74 (1979).
21. D. P. Peterson and D. Middleton, *Info. and Control* 5, 279 (1962).
22. R. Bernstein, *IBM J. Res. Devel.*, p. 40 (1976).

Appendix II

Optical/Digital Hybrid System Architectures for Interframe (Temporal) Image Data Compression

Optical/Digital Hybrid System Architectures
for Interframe (Temporal) Image Data Compression

Research Objectives

Interframe data compression with optical computations has been demonstrated under previous support of AFOSR. Interframe data compression has proven more elusive, principally because of the complex logic operations required in interframe comparisons. The objective of this research is to demonstrate that architectures for interframe compression can be conceived which combine computations appropriate with digital processors.

Appendix Contents

The contents of this appendix consists of a review of the general problem of interframe compression plus a proposed architecture for an optical/digital system for same, plus a paper demonstrating the results achieved in simulations of one such method.

I. INTRODUCTION

A hybrid optical/digital system architecture has been shown in previous research to be a feasible and viable approach to interframe coder design [1,2]. One of the definite advantages of optical computation over digital computation is the speed and ease of frame differencing in terms of the inherent image-plane-to-image-plane parallel processing. The temporally adaptive hybrid optical/digital system proposed in Ref. [2] was designed to use the above advantage, as well as the temporal redundancy between successive image sequences. However, it is not quite spatially adaptive due to the fact that only the interpolated differential pulse code modulation (IDPCM) spatial compression system has been used where large movement greater than the predetermined threshold is involved between image sequences. Therefore, the above scheme can be greatly improved by considering the following factors:

- (1) An adaptively-coordinated spatial/temporal subsampling and interpolation between the spatial and temporal compression subsystems are feasible by using programmable resolution provided by focal plane detector technology.
- (2) Motion-detection/estimation schemes could be integrated into a hybrid optical/digital architecture.
- (3) More reliable movement activity measures (image derivative energy) [3] could replace the normalized mean square error (NMSE) of the frame difference, which could be used for the decision mechanism of an adaptive compression system.

- (4) A highly parallel digital processing scheme using multiple microprocessors could be integrated into the hybrid optical/digital architecture in order to offset the slow processing speed of sequential digital subsystems as opposed to the high-speed optical subsystem. Also, the problem of efficient interfacing between the optical and digital subsystems should not be overlooked.

This paper addresses the above suggestions for the improvement of the hybrid optical/digital approach for interframe image data compression by introducing the adaptive decision mechanism based on image derivative energy. In Chapter II, the state of the art of the adaptive digital interframe predictive coding is reviewed. In Chapter III, we propose the use of image activity measurement for adaptive multimode coding, which is designed to integrate high-speed optical processing for the computation of spatial and temporal gradients and energies, and flexible digital processing for controlling multimode coding. In addition to the system configuration and algorithmic explanation, the current technological trends are briefly described to justify parallel processing and hybrid optical/digital processing.

Finally, the striking similarity between the parallel image processor in the proposed adaptive multimode coder and the parallel processor PASM (partitionable SIMD/MIMD system) has been pointed out. Thus, it is inferred that the parallel hybrid optical/digital processing approach provides a much broader conceptual framework for computer vision systems. Detailed architectural consideration will be left for further research.

II. ADAPTIVE DIGITAL INTERFRAME PREDICTIVE CODING: A REVIEW

There have been numerous excellent surveys concerning digital picture coding [4-9]. Rather than repeating the above surveys, we selectively consider the recent development related to adaptive interframe predictive coding.

There are various adaptive strategies that have been developed for spatial compression algorithms, such as differential pulse code modulation (DPCM), etc. The typical examples are adaptive prediction and quantization approaches. The former is based on the nonstationary statistical image model and assumption, which are more reliable in characterizing the edges or outline of the real world images than the first order Markov image model. Thus, the prediction error at the edge between the adaptive predictor and the original pixel is made smaller than the nonadaptive prediction error. The latter takes advantage of the insensitivity of the human eye to the quantization error. That is, the quantization steps and levels are adaptively selected in such a way that the introduced quantization distortion is not perceptable to the human visual system.

Interframe compression algorithms also take advantage of the temporal characteristics of the human visual system, in addition to the above spatial characteristics of human visual perception and the statistical redundancy represented by an image source model. A recent study [9] on the spatio-temporal response of the human visual system revealed that, at high temporal frequencies, spatial contrast sensitivity is reduced. This indicates that the sensitivity of the human visual system is low to fine spatial

detail in rapidly-moving objects, and at high spatial frequencies there is an overall decrease in flicker sensitivity. For the nontracking task, which is typical in teleconference or picture phone environments, the human visual system can tolerate a loss of spatial resolution in the moving area (reported by Miyahara [10]). These observations strongly suggest that the interframe coder can classify the parts of the spatial source image sequences and adaptively apply different efficient coding schemes to them, in accordance with the different spatial and temporal resolutions required by the human viewer for the appropriate visual perception. The simplest coding schemes are the frame replenishment method using frame repeating-dropping, spatial and temporal subsampling and interpolation.

The more sophisticated methods consist of extracting the actual motion displacement between frames. This motion estimation approach is still in its infancy and most of the past work has been restricted to two-dimensional motion, especially translation [9,11]. There have been four major approaches to estimating two-dimensional translation: (1) the Fourier method, (2) the method of differencing, (3) the temporal-spatial gradient technique, and (4) matching. In the context of TV signal transmission, the latter three have been selected for real-time implementation. The basis of movement-compensated predictive coding is that, if the displacement field of the moving-image sequence is known, then a very good prediction of the current field in the moving image sequence can be obtained by shifting and interpolating the parts of the previous field (or frame) which have moved [9]. This general approach is also simplified into two methods for real-time implementation. One method is the Pel-block displacement

estimation/compensation technique [12-18] and the other is the Pel-recursive displacement estimation/compensation technique [19,20,25].

In the former, the scene is divided into rectangular blocks, and a single shift is estimated for each block. The block may or may not be segmented into changed and stationary areas. The previous frame (or field) is shifted, and interpolated if nonintegral displacement estimates are allowed, to form the prediction for that block [9].

The latter technique updates the displacement estimate (which may or may not be obtained by Pel-block motion estimation) at each pixel. This update is based only on previously transmitted pels, so that no explicit displacement estimate need be transmitted.

There are two methods for estimating translational displacements: (1) the correlation or matching technique [12,13,18], and (2) the temporal-spatial gradient technique [10,14,17,21,22,23]. Given an object in translational motion with velocity $\underline{V}=(V_1, V_2)$, the image luminance satisfies

$$U(\underline{X}, t) = U[\underline{X} - \underline{V}(t - t_0), t_0] \quad (\text{II.1})$$

for an arbitrary reference time t_0 . In particular, if τ is the time interval between frames

$$U(\underline{X}, t_0 + \tau) = U[\underline{X} - \underline{V}\tau, t_0], \quad (\text{II.2})$$

where $\underline{d} = \tau \underline{V}$ is the displacement which occurs in one frame interval.

II.1. The Correlation/Matching Technique

The basic idea of this method is to choose \underline{d} from a set of possible displacements \mathcal{D} to minimize some measure of the difference between the current frame and the previous frame displaced by \underline{d} . Thus, $\hat{\underline{d}}$ is the value which solves

$$\min_{\underline{d} \in \mathcal{D}} \sum_{\underline{x} \in \text{MA}} N[U(\underline{x}, t_0 + \tau) - U(\underline{x} - \underline{d}, t_0)] \quad (\text{II.3})$$

where MA denotes the moving area and $N(\cdot)$ is a nonnegative increasing function which serves as a distance measure.

Equation (II.3) indicates that a large computation is required, since it should be evaluated for each $\underline{d} \in \mathcal{D}$. Thus, there is a trade-off between accuracy and computational load. In practice, only a relatively small number of integral displacements have been used.

II.2. The Temporal-Spatial Gradient Technique [14,21,22,23]

The standard procedure to minimize the difference between $U(\underline{x}, t)$ and $\hat{U}(\underline{x}, t) = U[\underline{x} - \underline{v}(t - t_0), t_0]$ is to take the derivative of the squared error $\varepsilon = [U(\underline{x}, t) - \hat{U}(\underline{x}, t)]^2$ and equate it to zero. Since there exists a relation between the temporal derivative and the spatial gradient of the moving object as shown in Ref. [24]:

$$\frac{\partial U(\underline{x}, t)}{\partial t} = -\underline{v} \cdot \nabla_{\underline{x}} U(\underline{x}, t), \quad (\text{II.4})$$

approximating the derivatives by finite differences, we have

$$\Delta T(\underline{X}, t) = - \underline{V} \cdot \Delta \underline{X}(\underline{X}, t). \quad (\text{II.5})$$

Thus, the velocity estimate is obtained by choosing \underline{V} , which gives the best mean-square fit to (II.5) over the moving area, i.e., which minimizes

$$C = \sum_{\underline{X} \in \text{MA}} [\Delta T(\underline{X}, t) + \underline{V} \cdot \Delta \underline{X}(\underline{X}, t)]^2. \quad (\text{II.6})$$

By setting $\Delta_V C = 0$, we obtain

$$\hat{\underline{V}} = - \left[\sum_{\underline{X} \in \text{MA}} \Delta \underline{X}(\underline{X}, t) \Delta \underline{X}(\underline{X}, t)^T \right]^{-1} \sum_{\underline{X} \in \text{MA}} \Delta T(\underline{X}, t) \Delta \underline{X}(\underline{X}, t). \quad (\text{II.7})$$

Limb and Murphy [10,17] have described a simplified version of the above estimator as follows:

$$\hat{\underline{V}} = - \begin{bmatrix} \sum_{\underline{X} \in \text{MA}} \Delta T(\underline{X}, t) \text{sign}[\Delta X_1(\underline{X}, t)] / \sum_{\underline{X} \in \text{MA}} |\Delta X_1(\underline{X}, t)| \\ \sum_{\underline{X} \in \text{MA}} \Delta T(\underline{X}, t) \text{sign}[\Delta X_2(\underline{X}, t)] / \sum_{\underline{X} \in \text{MA}} |\Delta X_2(\underline{X}, t)| \end{bmatrix} \quad (\text{II.8})$$

where $\underline{X} = (X_1, X_2)$.

A survey of these two methods is given in Ref. [11]. The gradient methods in general require fewer computations than the matching/correlation technique for a given level of accuracy. This is particularly critical for real-time application in image sequence (interframe) coding. However, the gradient method is inaccurate for large displacements (above several pixels per frame) where there is a significant amount of detail due to the fact that the temporal derivative in Eq. (II.4) is approximated by the frame difference. This problem can be alleviated by introducing a temporal recursion. In other words, the corresponding estimate for the previous frame is updated using the same algorithm with the Pel-Block displacement estimation technique, but the frame difference is replaced by the displaced frame difference. The resulting estimator, called the directional gradient method or the Pel-recursive displacement estimation technique, is accurate and robust to noise for large area translation, but it cannot respond to rapid spatial changes in the displacement field.

Netravali and Robbins [19] have proposed a coder in which the displacement estimate is recursively updated at a spatially-neighboring point as follows: Given the displaced frame difference, with displacement \underline{d} ,

$$D(\underline{X}, t, \underline{d}) = U(\underline{X}, t) - U(\underline{X} - \underline{d}, t - \tau), \quad (\text{II.9})$$

and if the object has translated by an amount \underline{d} in one frame, then $D(\underline{X}, t, \underline{d}) = 0$. The basis of the pel-recursive estimator is to adjust $\hat{\underline{d}}$ at each pel in order to reduce $|D(\underline{X}, t, \hat{\underline{d}})|$. This is done by changing $\hat{\underline{d}}$ in the direction of the negative gradient, that is,

$$\underline{\hat{d}}^i = \underline{\hat{d}}^{i-1} - \frac{\epsilon}{2} \nabla_d [D(\underline{X}, t, \underline{\hat{d}}^{i-1})]^2 \quad (\text{II.10})$$

where $\underline{\hat{d}}^i$ is the displacement estimate at pel i of some scanning sequence. Inserting Eq. (II.9) into Eq. (II.10) yields

$$\underline{\hat{d}}^i = \underline{\hat{d}}^{i-1} - \epsilon D(\underline{X}, t, \underline{\hat{d}}^{i-1}) \nabla_{\underline{X}} U(\underline{X} - \underline{\hat{d}}^{i-1}, t - \tau). \quad (\text{II.11})$$

Since $U(\underline{X}, t)$ is defined only at the sampling point, interpolation must be used to evaluate $D(\underline{X}, t, \underline{\hat{d}})$ and $\nabla_{\underline{X}} U(\underline{X} - \underline{\hat{d}}, t - \tau)$. They found it adequate to use linear interpolation to evaluate $U(\underline{X} - \underline{\hat{d}}, t - \tau)$, and $\nabla_{\underline{X}} U(\underline{X} - [\underline{\hat{d}}], t - \tau)$ to estimate the gradient, where $[\underline{\hat{d}}]$ is $\underline{\hat{d}}$ rounded to the nearest grid point. This algorithm, where $\epsilon = 1/1024$, further reduced the bit rate by 1 to 1.5 bits per pixel for the displaced-frame difference with respect to simple frame difference. However, the Pel-recursive algorithm has limited spatial convergence to ensure stability and is more sensitive to noise.

The Pel-recursive algorithm is also accompanied by the problem of addressing the moving and stationary areas and coding the movement-compensated prediction error in moving areas. Upon successful movement compensation, most of the quantized prediction errors will be zero, thus, use of efficient run-length coding is justified. Also, since there are three states (stationary area, insignificant compensated prediction errors, and significant compensated prediction errors), an identification for the state of the next run length must be provided in addition to the run lengths. An

alternative to this three-state encoding is to run-length code only significant and insignificant prediction errors. The decision as to which prediction to use (i.e., previous frame or displaced previous frame) can be made by using previously transmitted information, so that no over head information need be transmitted.

The Pel-recursive displacement estimation has been extended to the transform domain in Ref. [25]. The algorithm works recursively in such a way that the displacement estimates are updated from two-dimensional unitary transform coefficients to two-dimensional unitary transform coefficients. Each image is partitioned into blocks of size $N=N_L \times N_C$, to obtain the N transform coefficients for each q th block subimage,

$$\begin{aligned} C_n(q) &= U^T(\underline{X}_q, t) \phi_n \\ \underline{X}_q &= (X_{1q}, X_{2q})^T \end{aligned} \quad (II.12)$$

where

$\underline{X}_q = (X_{1q}, X_{2q})^T$ denote the coordinates of the upper left-hand Pel of the q th block

$U(\underline{X}_q, t)$ = the column scanned vector of intensities of block q

$C_n(q)$ = the n th transform coefficient

ϕ_n = the n th linear transform basis vector.

Similarly,

$$C_n(q, \hat{D}) = U^T(\underline{X}_q - \hat{D}, t - \tau) \phi_n \quad (II.13)$$

c is the n th coefficient in the previous displaced frame. When $C_n(q)$ is predicted by $C_n(q, \hat{D})$, the prediction error is

$$C(q, \hat{D}) = C_n(q) - C_n(q, \hat{D}). \quad (II.14)$$

The squared prediction errors are minimized by a steepest descent method. The resulting recursive formula is

$$\hat{D}_{n+1}(q) = \hat{D}(q) - \frac{\epsilon}{2} \nabla_{D_n(q)} e_n^2(q, \hat{D}_n(q)).$$

Jones and Rashid [20] have proposed residual-recursive displacement estimation, which utilizes adaptive estimation techniques to vary certain parameters of adaptive hybrid picture coding (AHPC) as the statistics of the image change. In AHPC, a one-dimensionally transformed $S(k)$ in a raster scan fashion is predicted by

$$S(k) = \sum_{i=1}^P a_i(k) S(k-i) + e(k), \quad (II.15)$$

where the predictor coefficient $a_i(k)$ is a piece-wise constant. If $a_i(k)$ is estimated, the residual sequence elements are given by

$$e(k) = S(k) - \sum_{i=1}^P \hat{a}_i(k) S(k-i), \quad (II.16)$$

where $e(k)$ is the difference between the prediction of the transformed picture element and the actual value. Assume that the image is divided into

blocks of N_r rows by N_c columns. Then each residual in the present frame may be represented by

$$e_n(q) = U^T(\underline{X}_q, t) \phi_n - U^T(\underline{X}_q, t) \phi_n, \quad (\text{II.17})$$

where $\overline{U^T(\underline{X}_q, t) \phi_n}$ is a linear prediction.

Similarly, the displaced previous frame value for this residual is

$$\hat{e}_n(q, \hat{D}) = U^T(\underline{X}_q - \hat{D}, t - \tau) \phi_n - U^T(\underline{X}_q - \hat{D}, t - \tau) \phi_n. \quad (\text{II.18})$$

A displaced residual difference may be formed by

$$\text{DRD}_n(q, \hat{D}) = e_n(q) - \hat{e}_n(q, \hat{D}). \quad (\text{II.19})$$

The algorithm attempts to minimize the squared displaced residual difference in a residual recursive steepest descent procedure as follows:

$$\hat{D}_{n+1}(q) = \hat{D}_n(q) - \frac{\epsilon}{2} \nabla_{\hat{D}_n(q)} [\text{DRD}_n(q, \hat{D})]^2. \quad (\text{II.20})$$

This operation may be simplified by

$$\hat{D}_{n+1}(q) = \hat{D}_n(q) - \epsilon \text{DRD}(q, \hat{D}(q)) e_n(\underline{X}_q - \hat{D}, T - \tau). \quad (\text{II.21})$$

They have shown that this algorithm is more stable in the region $\epsilon = 10^{-4}$ than the coefficient recursive displacement estimation algorithm.

III. A SPATIO-TEMPORALLY ADAPTIVE MULTIMODE CODING UNDER PARALLEL HYBRID OPTICAL/DIGITAL PROCESSING CONCEPT

III.1. Adaptive Multimode Coding Using Image Activity Measure

As discussed in the previous chapter, the state-of-the-art interframe coding algorithms have become very sophisticated at the cost of implementation and system complexity. However, the compressed bit rates remain beyond the capabilities of conventional digital transmission channels, such as telephone links. The common characteristic of most interframe coding implementation is the need for a frame buffer for storing and smoothing the image information to be transmitted. Since the size of the buffer is limited by constraints relating to cost and perceptual delay effects, peaks of activity in the image signal can cause this buffer to overflow. To avoid this, it is necessary to adapt the coder operation to the amount of motion, that is, to deliberately degrade the image quality in a gradual and graceful manner as the image activity increases. Depending on the desired transmission rate and image quality, it may be possible, under the conditions of moderate motion, to utilize various psychovisual properties described in the previous chapter, such as spatio-temporal response, so that the adaptive strategies do not introduce visible degradations. Nevertheless, for violent motion, visible degradation may be unavoidable. Therefore, consideration should be given as to what extent visible degradations can be tolerated by a human viewer for the application in mind.

The ideal adaptive strategies should depend on the actual characteristics of the image signal which produce the nonuniform information rate, such as percentage of frame motion, velocity of motion, amount of

spatial detail, etc. However, most multimode coders use a single decision parameter, i.e., the buffer memory occupancy rate, which has also been used under hybrid optical/digital compression schemes (see Ref. [2]). This has lead to some degree of success in real-time multimode interframe codec simulations for coding National Television System Committee (NTSC) monochrome video signals at 1.5 Mbits/sec under minicomputer control [26-28]. These can be called a class of multimode coders where M modes of coding operations, such as were described in the previous chapter, are selectively applied to a video signal to assure its continuous operation, to provide the full available resolution in the lower mode, and to introduce the graceful and progressive degradation of image quality in the higher mode. The basic design problem of a multimode coder is lack of a "genuine" adaptive mechanism to switch between the higher and lower mode of operations. Instead of using a buffer memory occupancy criterion, we propose the use of image activity measures. One measure of image activity that can be readily computed by an optical processor is image derivative energy for spatial and temporal gradients, as proposed by McCaughey [3].

III.2. Parallel Hybrid Optical/Digital Processing Approach To Interframe Coding

The use of optical processing for image data compression has been oriented toward overcoming the excessive digital processing requirements of digital image bandwidth compression techniques. The advantages of optical processing of images are evident in light of current and emerging device technologies such as focal plane CCD or CID detectors, spatial light modulators (PROM, LCLV, high-speed silicon lithium niobate (SiLiNb) devices, etc.), multi-mode optical fibers coupled with laser-based optical communication technology, and VLSI technology enabling the potential realization of parallel digital image processors. Within this new technological environment, it is fruitful to reevaluate the advantages of both optical and digital processing for various real-time applications, such as image data compression for TV conferencing, military and/or medical problems based on multiple image sequence, etc. The overall goal is system architecture design, i.e., combining advanced optical and digital technologies to improve the performance of compression systems.

The viable guideline for hybrid optical/digital system design would be a full use of parallelism in image processing. The conventional and current image transmission systems have been restricted by a sequential, or raster scanning mode of operation. There is good reason for it in terms of simple, real-time implementation. However, it may be useful to reconsider the available modes of operation, from sequential line-by-line to parallel block-by-block in spatial domain.

Returning to motion-compensated coding algorithms, which are the most sophisticated that have been developed thus far, it has been noted that they

suffer from excessive search computation and stability problems. However, not all the image subblocks need motion compensation. Also, the spatial and temporal gradient information needed for the motion estimation can be easily computed as a low-end operation by the smart sensor arrays of an optical processor. The image spatial/temporal derivative energies which are used as a criteria for selecting an appropriate mode are also readily computed by an optical processor. Finally, the hypothetical multimode coder, consisting of N subimage buffers and M microprocessors, can be viewed as a single-instruction multiple-data stream/multiple-instruction multiple-data stream (SIMD/MIMD) parallel processor. Viewed in this way, the adaptive decision mechanism can be considered a special-purpose computer (software and/or hardware) interfaced with both optical and digital processors, which takes information from the low-end optical subsystem monitoring image activities, processes them based on the preprogrammed decision rules, sends out the instructions to the digital subsystem for multimode coding as to which image subblock image buffer needs a particular mode of processing by a particular microprocessor, and generates the overhead information to be transmitted to the receiver.

The details of the proposed adaptive multimode predictive coding based on parallel image processing concepts are discussed next.

III.3. System Configuration of Adaptive Multimode Coding

The basic structure that we propose for the hybrid optical/digital adaptive multimode coder is given in Fig. 1. The image sequence is captured by a "smart sensor" optical preprocessor that generates the spatial and temporal gradients for each subimage block. The spatial and temporal image derivative energies to each subblock can be computed by either a digital or an optical preprocessor. Now the spatial and temporal image derivative energy values for each subblock are used for adaptive selection of the coding mode applied to that subblock or the temporal derivative and spatial gradient values at each pel may be used to implement Pel-recursive motion compensation coding on a sequential basis as well.

The adaptive decision mechanism thresholds the image activity measure values for each block and generates the "instructions" as to what mode of operation instruction (software program) is loaded into which microprocessor and which subimage memory content is fetched for execution. When the Pel-block motion compensation algorithm is applied to the particular subimage, the overhead information about whether the subimage block is the stationary or moving part of the image must be sent back to the adaptive decision mechanism. If the Pel-recursive algorithm is applied, the address information of the moving pel should be generated on a sequential basis. The adaptive decision mechanism can access the frame buffer for smoothing the data rate in order to double-check its performance. Therefore, when combined with the overhead information generator, it is a solely digital subsystem as is, for example, a conventional minicomputer.

The parallel image processor consists basically of N subblocks, image/frame buffers/memories and M microprocessors. Thus, it is also a

digital subsystem. M microprocessors for image processing can perform P different modes of image manipulations. There are various operations that can be used for these P modes, such as 8-bit PCM, spatial/temporal subsampling, frame repeating/dropping, motion compensation coding, adaptive quantization, linear and nonlinear temporal filtering, etc. Given the image processing modes, intensive simulation experiments with subjective viewing should be used to find the appropriate threshold values for the image activity measures that can be used for mode switching. In our subsequent simulation experiments, five modes of operation are used: (1) 8-bit PCM, (2) spatial subsampling (subsampling every two pixels), (3) frame repeating, (4) motion-compensation, and (5) temporal filtering.

Finally, the postprocessor or reconstruction processor at the receiving end is naturally conceived as a digital subsystem that combines the transmitted pieces of subimage blocks and treats the possible block discontinuity in the reconstructed image with the aid of the transmitted overhead information for subsequent display.

III.4. Algorithm of the Proposed Adaptive Multimode Coder

- Step 1. Divide the image fields $U(\underline{X}, t)$ and $U(\underline{X}, t+\tau)$ into small subblocks (say, 8×8). Thus, there are 1024 8×8 subblocks for a 256×256 image.
- Step 2. Compute the temporal differences of the corresponding i th subblock between fields $U(\underline{X}, t)$ and $U(\underline{X}, t+\tau)$, such as $U_{T_i}(\tau) = |U_i(\underline{X}, t+\tau) - U_i(\underline{X}, t)|$ for $i = 1, \dots, N = 1024$.
- Step 3. Compute the temporal image derivative energy for the i th temporal subblock by using the following formula:

$$FE_i \triangleq \left| \frac{d U_{T_i}(\tau)}{d\tau} \right|^2$$

for the i th subimage.

- Step 4. Compute the spatial image derivative energy for the i th spatial subblock by using the following formula:

$$SE_i \triangleq \left| \frac{\partial^2 U_i(\underline{X})}{\partial^2 \underline{X}} \right|^2 \quad \text{for the } i\text{th subimage where } \underline{X} = (X_1, X_2)$$

$$= [U_i(X_1-1, X_2) - 2U_i(X_1, X_2) + U_i(X_1+1, X_2)]^2$$

$$+ [U_i(X_1, X_2-1) - 2U_i(X_1, X_2) + U_i(X_1, X_2+1)]^2.$$

- Step 5. Adaptive decision-making based on thresholding takes place for each subblock, $i=1,\dots,N$ (see Table 1). This rule may be changed, depending on the subjective evaluation tests.
- Step 6. Generate the overhead information for each subblock $i = 1,\dots,N$; in other words, information about which subblock image i is processed by a particular operation mode (1) frame repeat, (2) spatial subsampling, (3) motion compensation, (4) temporal filtering, (5) 8-bit PCM, and the address of the moving and stationary areas in the i th subblock, etc.
- Step 7. Reconstruct the block-by-block transmitted mosaic image by using the overhead information and, if necessary, cosmetically treat the block discontinuity by spatial filtering or contrast adjustment.

IV. PARALLEL IMAGE PROCESSING ARCHITECTURE FOR MULTIMODE CODING: FURTHER RESEARCH AREAS

There are some remaining questions in the parallel image processing approach to multimode coding. They are concerned with how to efficiently interconnect the subimage buffers/memories with the multimicroprocessors. The interconnection network problem for parallel and distributed processing has been intensively researched by the computer architecture community during the past several years. Siegel et al. [29,30] have proposed PASM: a partitionable SIMD/MIMD system for image processing and pattern recognition. Figure 2 shows the block diagram overview of PASM. There is a striking resemblance between Figs. 1 and 2, except for the existence of the overhead information generator, frame buffer and transmission channel. In other words, the parallel digital image processor in Fig. 1 seems to encompass the memory storage system, the memory management system, the parallel computation unit, and the microcontrollers in Fig. 2. These units are controlled by the system control unit, and correspond to the digital decision processor in Fig. 1. Therefore, the same kind of considerations toward PASM architecture design seem to apply to the multimode coder design as well.

It seems that the conceptual framework of parallel digital image processing is not restricted to image transmission, but can also apply more generally to an intelligent computer vision application. However, optical computation is an essential part of preprocessing used in the computation of the spatio-temporal gradients and energies in the proposed interframe compression scheme. It has also been raised to the sophisticated level of

digital computation, provided that parallel hybrid optical/digital systems architecture is practically implementable.

In the near future, the performance of the proposed adaptive multimode coding algorithm and its channel error effects will be evaluated by simulation experiments done on a PDP 11/70 and on an I²S image processor. Also, if time permits, conceptual architecture of a parallel digital image processor for a multimode coder will be compared with PASM architecture.

REFERENCES

1. Hunt, B. R. and Ito, H. N., "A hybrid optical/digital interframe image data compression scheme," Proc. SPIE 249, 1980.
2. Ito, H. N. and Hunt, B. R. , "A temporally adaptive hybrid optical/digital interframe compression scheme," Proc. SPIE 292, 1981.
3. McCaughey, D. G., "Variable resolution hybrid-coding techniques," Acoust. Speech and Signal Processing Workshop on Two-Dimensional Signal Processing, Berkeley, California, 1979.
4. Pratt, W. K., ed., Image Transmission Technique, (Academic Press, New York, N. Y., 1979).
5. Netravalli, A. N. and Limb, J. O., "Picture coding: A review," Proc. IEEE 68:306, 1980.
6. Jain, A. K., "Image data compression: A review," Proc. IEEE 69:349, 1981.
7. Stafford, R. H., Digital Television: Bandwidth Reduction and Communication Aspects, (John-Wiley and Sons, New York, N. Y., 1980).
8. Haskell, B. G. and Steele, R., "Audio and video bit-rate reduction," Proc. IEEE 69:252, 1981.
9. Huang, T. S., Image Sequence Analysis, (Springer-Verlag, Berlin, 1981).
10. Miyahara, M., "Analysis of perception of motion in television signals and its application to bandwidth compression," IEEE Trans. Comm. COM-23:761, 1975.
11. Thompson, W. B. and Barnard, S. T., "Lower-level estimation and interpretation of visual motion," Computer 14:20, 1981.

12. Rocca, F. and Zanoletti, S., "Bandwidth reduction via movement compensation on a model of the random video process," IEEE Trans. Comm. :960, 1972.
13. Giorda, F. and Racci, A., "Bandwidth reduction of video signals via shift vector transmission," IEEE Trans. Comm., p. 1002, 1975.
14. Cafforio, C. and Rocca, F., "Methods for measuring small displacements of television images," IEEE Trans. Info Th. IT-22:573, 1976.
15. Brofforio, S. and Rocca, F., "Interframe redundancy reduction of video signals generated by translating objects," IEEE Trans. Comm, COM-25:448, 1977.
16. Limb, J. O. and Murphy, J. A., "Measuring the speed of moving objects from television signals," IEEE Trans. Comm. COM-23:474, 1975.
17. Limb, J. O. and Murphy, J. A., "Estimating the velocity of moving images in television signals," Computer Graphics and Image Processing 4:311, 1975.
18. Jain, J. R. and Jain, A. K., "Displacement measurement and its application in interframe image coding," IEEE Trans. Comm. COM-29:1799, 1981.
19. Netravali, A. N. and Robbins, J. D., "Motion-compensated television coding: Part I," Bell Syst. Tech. J., 58:631, 1979.
20. Jones, R. A. and Rashid, H. V., "Residual recursive displacement estimation," Proc. PRIP , p. 508, 1981.
21. Paquin, R. and Dubois, E., "Estimation of the displacement field of moving images," 1981 Picture Coding Symposium, Montreal, Canada, June 1981.
22. Dubois, E. "A multimode interframe coder using movement compensation," 1981 Picture Coding Symposium, Montreal, Canada, June 1981.

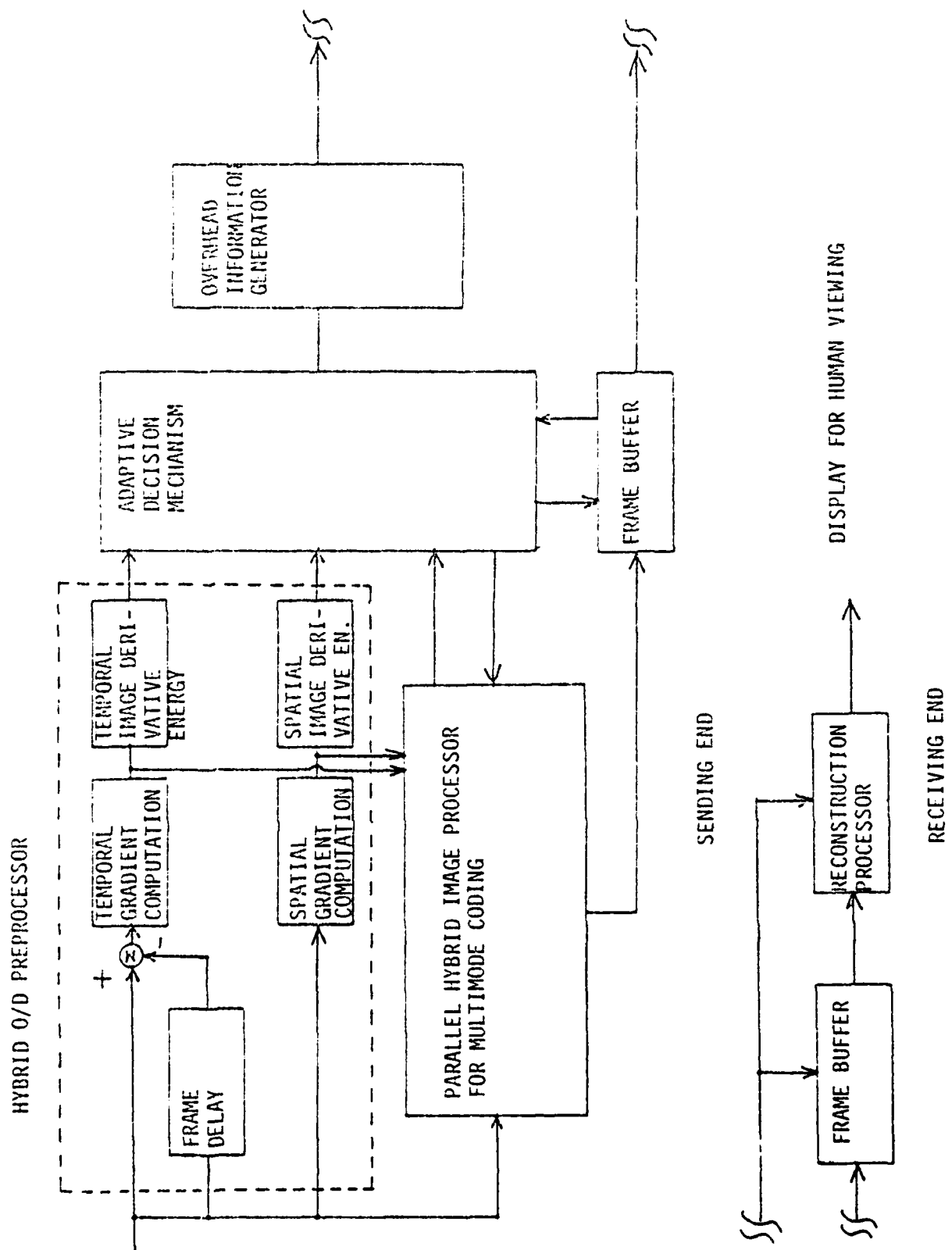
23. Paquin, R. and Dubois, E., "A spatio-temporal gradient method for estimating the displacement field in time-varying imagery," INRS Telecommunications, February 1982.
24. Schunck, B. G. and Horn, B. K. P., "Constraints on optical flow computation," Proc. PRIP, p. 205, 1981.
25. Stuller, J. A. and Netravali, A. N., "Transform domain motion estimation," BSTJ 58:1673, 1979.
26. Haskell, B. G., Gordon, P. L., Schmidt, R. L., and Scattaglia, J. V., "Interframe coding of 525-line monochrome television at 1.5 Mbits/sec," IEEE Trans. Comm. COM-25:1339, 1977.
27. Ishiguro, I., Iinuma, K., Iijima, Y., Koga, T., Azami, S., and Mune, T., "Composite interframe coding of NTSC color television signals," Proc. Nat. Telecomm. Conf., Dallas, Texas, 1976, Vol. I, pp 6.4-1 to 6.4-5
28. Yasuda, H., Kuroda, H., Kawanishi, H., Kanaya, F., and Hashimoto, H., "Transmitting 4 MHz TV signals by combinational difference coding," IEEE Trans. Comm. COM-25:508, 1977.
29. Siegel, H. J., Siegel, L. J., Kemmerer, F. C., Mueller, P. T., Jr., Smalley, H. E., Jr., and Smith, S. D., "PASM: A partitionable SIMD/MIMD system for image processing and pattern recognition," IEEE Trans. Comp. C-30:934, 1981.
30. Siegel, L. J., Siegel, H. J., and Feather, A. E., "Parallel processing approaches to image correlation," IEEE Trans. Comp. C-31:208, 1982.

LIST OF FIGURES

Fig. 1. System configuration of adaptive multimode coding.

Fig. 2. Block diagram overview of PASM.

Fig. 1.
to



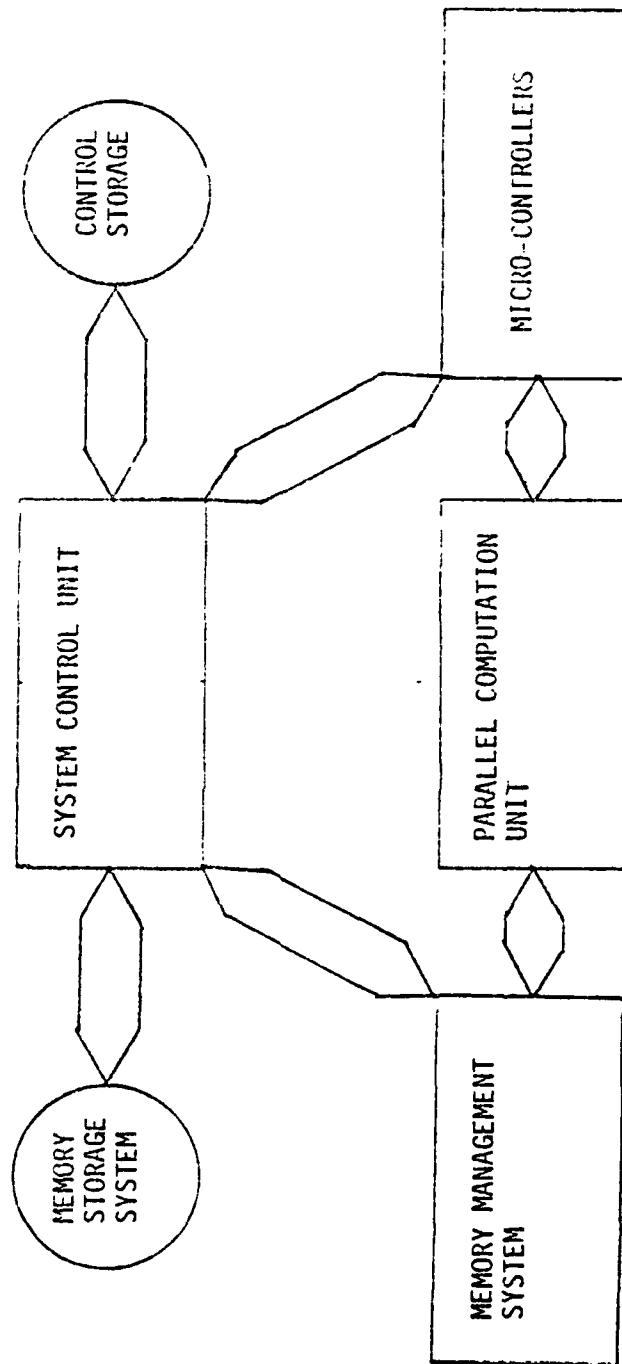


Fig. 2.
10

Table 1. Adaptive decision mechanism.

TEMPORAL IMAGE ACTIVITY MEASURE FE_{Bi}	SPATIAL IMAGE ACTIVITY MEASURE SE_{Bi}			
	$SE_{Bi} < T_{S1}$	$T_{S1} \leq SE_{Bi} \leq T_{S2}$	$T_{S2} < SE_{Bi}$	
	FRAME REPEAT	SPATIAL SUBSAMPLING OR ELEMENT DPCM	SPATIAL SUBSAMPLING OR ELEMENT DPCM	
	MOTION COMPENSATION	MOTION COMPENSATION	MOTION COMPENSATION	
	$FE_{Bi} < T_{F1}$			
	$T_{F1} \leq FE_{Bi} \leq T_{F2}$			
	$T_{F2} < FE_{Bi}$			
		TEMPORAL FILTERING OR ELEMENT DPCM	TEMPORAL FILTERING OR ELEMENT DPCM	8 BIT PCM

SPIE Conference August, 1981 San Diego

A Temporally Adaptive Hybrid Optical/Digital Interframe Compression Scheme

H. N. Ito and B. R. Hunt

Digital Image Analysis Laboratory
 Department of Systems Engineering
 University of Arizona
 Tucson, Arizona 85721

Abstract

The feasibility of optical processing for interframe image compression has been demonstrated by hybrid optical/digital IDPCM/Frame-to-Frame DPCM interframe compression architecture and its simulation [1]. However, the subjective image quality of the reconstructed 14th frame for 1.5 bits/pixel or 1.75 bits/pixel are not excellent enough to meet a broadcasting standard. Also, the transmission rate performances per second with the frame rate: 30 frames/sec and 256 x 256 lines of the images are 1.9 Mbits/sec or 3.4 Mbits/sec, respectively, which are not quite competitive against the 1.5 Mbits/sec monochrome videotelephone system performance [2]. In this paper, in order to improve the hybrid system performance, another hybrid optical/digital adaptive IDPCM/conditional replenishment interframe compression architecture is proposed and simulated.

As a result of simulating 14 sequences of Walter Cronkite images with 256 x 256 lines and 0 bit intensity, reconstructed images with an excellent image quality are obtained with 1.72 Mbits/sec for the frame rate: 30 frames/sec, and the average compression ratio performance is 16.84. And the average D.T rate is 0.37 bits/pixel.

Introduction

Image data compression is a very active topic of research in image processing [3], [4]. Traditionally, most image data compression schemes are dominated by digital processing due to the fact that digital systems are inherently flexible and reliable. However, it has been demonstrated that optical processing can be used for image data compression, by Hunt, et.al. [5], [6], and [7]. One of the schemes for image data compression architecture with optical implementation is the IDPCM method [5]. This is a data compression scheme which functions analogously to conventional digital DPCM compression, except that the specific compression steps are implemented by incoherent optical processing. The extension of the IDPCM method into interframe compression using a frame-to-frame digital DPCM scheme leads to hybrid optical/digital architecture with electro-optical devices [7], [8], such as PDM or CCD or liquid crystal light valve.

The bottleneck of this proposed hybrid optical/digital interframe compression scheme is mainly caused by the motion displacement contained in neighboring frames as well as the frame prediction mechanism. The predicted frame used in [7] is just a previous frame, assuming the motion displacement between frames is small. This is not usually the case in real interframe image sequences. Therefore, motion displacement errors between neighboring frames are accumulated as time and frames go on, and unacceptable distortion is generated in the reconstructed image if the low frequency quantization level is not large enough to allow the low frequency component which constitutes most of the motion displacement to be correctly coded. Also, the subjective image quality of the reconstructed 14th frame for 1.5 bits/pixel or 1.75 bits/pixel is not excellent enough to meet a broadcasting standard. In addition, the transmission rate performances per second with the frame rate: 30 frames/sec and 256 pixels x 256 lines of the images are 1.9 Mbits/sec or 3.4 Mbits/sec, respectively. This is not quite competitive against the 1.5 Mbits/sec of monochrome videotelephone system performance [2]. In this paper, in order to improve the hybrid optical/digital system performance, another hybrid optical/digital adaptive IDPCM/conditional replenishment interframe compression architecture is proposed and simulated. The conventional conditional replenishment technique is based on the segmentation of the moving pixels from the stationary pixels, and only the significantly moving pixels are transmitted along with their address information. At the receiving end, the stored preceding image is repeated with the significantly moving pixels replaced by the transmitted moving pixels. The following are problems frequently associated with the conditional replenishment technique: 1) buffer overflow, and 2) address coding. They will not be discussed in detail, except that the threshold level of the motion detector is recursively set so that the buffer can store the largest number of moving pixels within the buffer capacity.

The essential part of the conditional replenishment technique, the motion detection scheme, can be divided into direct and indirect motion detection methods. The adaptive

292-10

strategy can be incorporated in the conditional replenishment technique by using various modes. The above motion detection and adaptive strategies will be discussed in the following section under the constraint of the hybrid optical/digital implementation.

In the simulation experiment, we examine the performances of four systems including the proposed adaptive hybrid optical/digital IDPCM/conditional replenishment system.

System configuration of a temporally adaptive hybrid optical digital interframe compression scheme

Interpolative DPCM spatial compression scheme

Interframe coding is the most efficient image coding system in terms of the image data reduction rate because spatial redundancy as well as temporal redundancy can be eliminated if the movement involved between neighboring frames is assumed to be small. As one of the efficient spatial compression systems, it has been demonstrated that the interpolative DPCM (IDPCM) method can reconstruct an image with excellent quality (MSE = 0.007%) with a bit rate of 3 bits/pixel, which is better than a conventional DPCM method. Figure 1(a) shows the incoherent optical system implementation of IDPCM compression and Figure 1(b) shows the digital implementation of IDPCM compression. The main differences between IDPCM and DPCM are as follows.

- (1) IDPCM does not assume causality of a pixel sequence in an image line, that is, DPCM is a pixel-to-pixel operation as opposed to IDPCM which takes a parallel image plane to-image plane operation.
- (2) the pixel estimate of DPCM is replaced with a neighborhood smoothing of the entire image plane, which is called a low frequency image in IDPCM.
- (3) the pixel difference of DPCM is replaced with the difference between the original image and the estimated or smoothed low frequency image, which is called a high frequency image in IDPCM.
- (4) another significant difference is a non-feedback structure in IDPCM as opposed to a feedback structure in DPCM. The feedback of a pixel estimate in DPCM is used to cancel the accumulated quantization error from the quantized difference. However, by using two-channel incoherent optical system implementation, the feedback structure can be eliminated without accumulating the quantization error. This is due to the fact that the visibility of the distortion at the receiver is reduced by the two-channel structure in which the high frequency image containing various distortion is added to the low frequency image.

The similarities between IDPCM and DPCM can be summarized in the following two points which lead to a common data compression strategy.

- (1) the low frequencies in the image can be predicted, removed from transmission by a difference step, and then repeated at the receiver.
- (2) the high frequencies, which are not eliminated by prediction and differencing, possess only a small amount of the total image energy and can be accurately coded with a small number of bits.

In Hybrid Optical/Digital Implementation, the subsampling and interpolation operation can be realized by CCD and the differencing operation can be implemented by an appropriate opto-electronic device such as a liquid crystal light valve or a PROM.

The spatial compression performances of the IDPCM system for every 2nd pixel and every 4th pixel subsampling and interpolation are given in table 1(a) and 1(b), when L_{TQ} and TQ levels are changed from 2 to 64. The overall optimal performance (MSE and bit rate per pixel) seems to be given by the every 2nd pixel subsampling and interpolation with $L_{TQ} = 16$, $TQ = 2$ and MSE = 0.00715%. This combination of quantization level parameters has been used for the subsequent adaptive interframe compression simulation.

Adaptive interframe compression schemes

Some of the simple but basic interframe coding schemes are Frame-to-Frame DPCM coding and a conditional replenishment coding. The former has been combined with the IDPCM spatial compression scheme and simulated and the results are given in [7]. The results show that unless the L_{TQ} level is large enough ($L_{TQ} = 8$ or 16), the accumulated motion misplacement errors among a sequence of frames causes severe distortion and artifacts in the reconstructed image of the 10th frame. Also the subjective image quality of the reconstructed image for high L_{TQ} levels ($L_{TQ} = 16$) is not excellent enough to meet

AD-A117 722

ARIZONA UNIV TUCSON DIGITAL IMAGE ANALYSIS LAB
OPTICAL COMPUTATIONS FOR IMAGE BANDWIDTH COMPRESSION.(U)
MAY 82 B R HUNT, R SCHOWENGERDT

F/G 9/2

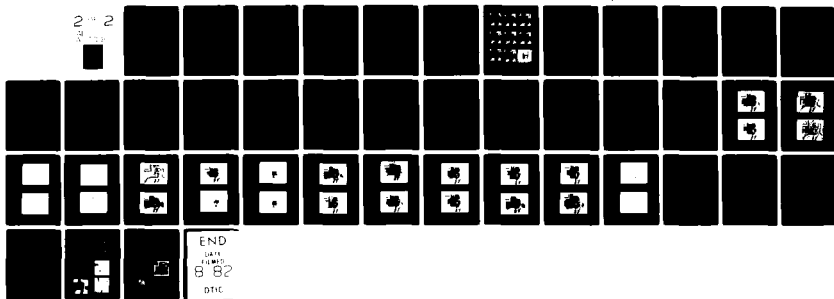
UNCLASSIFIED

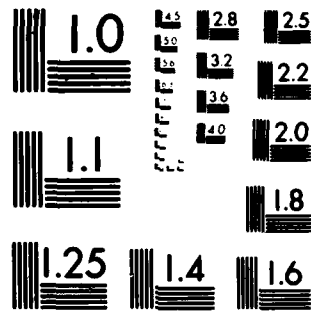
DIAL/EE-81-005

AFOSR-TR-82-05P0

AFOSR-81-0170

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

broadcasting standards, and the transmission rate per second with 30 frames/sec frame rate and 256 pixels x 256 lines of sequence of the images is 2.9-3.4 Mbits/sec with reasonably fair image quality.

It has been observed that the videoconference images involve mainly small movement so that the stationary pixels of the preceeding frame can be repeated to reconstruct the next frame at the receiver, and only the moving pixels are transmitted. In other words, the conditional replenishment coding can be expected to perform better than Frame-to-Frame DPCM coding in terms of both NMSE and transmission rate.

The typical system components of the conditional replenishment coding system are (1) motion detector or significant change detector, (2) pixel selector, based on the address information produced by (1), (3) uniform quantizer (64 levels), (4) address coder, (5) frame buffer, and (6) reconstructor.

The essential part of the conditional replenishment coder is the motion detector or significant change detector, which can be, in general, divided into two categories: direct and indirect methods. The direct method is based on laborious calculation of the velocity [9], [10] or the estimation of motion displacement of a moving pixel between neighboring frames, using a 3 dimensional image or field model [11], [12]. These methods are the bases for motion estimation/compensation coding whose significantly improved performance has been demonstrated for spatially uniform translated displacement by Netravali, et.al [13] and others. In order to implement conditional replenishment interframe coding combined with DPCM spatial compression within the framework of the hybrid optical/digital system, an indirect method of detecting significant change between neighboring frames by thresholding the frame difference is preferred to the direct method in which the velocity and/or motion displacement are calculated from the pixel-to-pixel operation. Also, the state of buffer to smooth moving pixel data generated by movement can be used as an indirect indication of the significant change. In addition, NMSE of the frame difference can indicate movement activity.

Within the framework of hybrid optical/digital architecture, the motion detector can be designed as a hybrid system, in which two liquid crystal light valves generate a frame difference image and the focal plane charge injection device (CID) converts the difference image intensity into a digitally addressable signal. Then, the decision making needed for the adaptive scheme can be made by using a microprocessor acting on the digital signal from the CID.

The adaptive strategies can be incorporated immediately into conditional replenishment coding, which is, indeed, a partially adaptive scheme itself, by accommodating several different modes. They are as follows:

- (1) Adaptive change of the threshold level depending on the movement activity and the buffer state.
- (2) Reduction of spatial and/or temporal resolution by using subsampling or filtering.
- (3) Suspension of replenishment due to large movement.
- (4) Field repeating due to very small or no movement.

Given the constraint of the optically implemented spatial compression system, the above adaptive strategy (1) has a limited utility because DPCM is not quite suitable for changing the spatial resolution of the limited region (moving area) in the image. Also, the image plane-to-image plane operation of DPCM makes it inefficient to directly calculate the velocity or motion displacement of significantly moving pixels as in digital processing. Therefore, we adopted an indirect change detection method using adaptive thresholding of the frame difference based on the buffer state and the adaptive strategy (3) the suspension of the replenishment, to form adaptive hybrid optical/digital DPCM/conditional replenishment interframe compression architecture as shown in Figure 2.

Simulation experiments and results

To demonstrate the superior performance of the adaptive hybrid optical/digital DPCM/conditional replenishment interframe compression architecture, a series of digital simulation experiments were carried out in the Digital Image Analysis Laboratory of the University of Arizona.

The source data for the simulation consisted of a sequence of 14 digitized frames from a television broadcast of Walter Cronkite. The frames are digitized at 256 x 256 pixels resolution, with 8 bits of intensity per pixel.

Assuming that address coding is accurate, no channel coding has been used, and no error has been added to the transmission channel, we have simulated the following four systems for comparison of the compression performances (bits/sec, with frame rate of 30 frames/sec) and image quality.

- (A) A nonadaptive conditional replenishment system with the 1st frame transmitted by 64 level PCM.
- (B) A nonadaptive hybrid optical/digital IDPCM/conditional replenishment system with the 1st frame transmitted by IDPCM (HQ = 2, LPUQ = 32).
- (C) An adaptive conditional replenishment system with 64 level PCM.
- (D) An adaptive hybrid optical/digital IDPCM/conditional replenishment system (HQ = 2, LPUQ = 32).

The adaptive strategies are based on frame replenishment suspension or frame repeating depending on the MSE of the successive frame difference. In other words, if MSE of the frame difference is less than 0.05%, the frame is repeated, and if MSE is more than 0.5%, the frame replenishment is suspended and only a spatially compressed image by 64 level PCM or IDPCM (LPUQ = 32, HQ = 2) is transmitted.

The motion detection strategy consists of frame differencing and iterative thresholding as shown in Figure 3, whose initial threshold is set as the mean value of the frame difference and one hundredth of the variance of the frame difference is incremented till the maximum number of moving pixels are detected within the maximum buffer size (3000 pixels) for uniform quantization.

Table 2 summarized the transmission rate (bits/sec), the average bit rate per frame (bits/frame) and the average compression per frame ratio, of four systems. Table 3 shows MSE, the number of moving pixels, transmitted, compression ratio, and the number of bits transmitted for the proposed adaptive hybrid optical/digital IDPCM/conditional replenishment system.

Graph 1 shows how the MSE performances of four systems change along 14 frames.

In the non-adaptive scheme, IDPCM based conditional replenishment system needs about a half of transmission channel capacity needed by PCM based one. And yet, the subjective and objective image qualities of the reconstructed pictures by both systems are unbearably impaired by a ghost type of noise or motion blur as shown in Picture 1.1 and 1.2. Also, as in the case of Hybrid optical/digital IDPCM/Frame-to-Frame DPCM coding system, the motion displacement between neighboring frames seems to be accumulated. This could be due to the fact that the buffer size is limited so that sufficient number of moving pixels are not correctly detected and transmitted to yield the reliable reconstruction at the receiving end.

In the adaptive scheme, in the same way as above, IDPCM based conditional replenishment system needs about one third of transmission channel capacity needed by PCM based one. Due to the adaptive strategy: Replenishment suspension taken place at the 7th frame till the 10th frame and the 14th frame, the subjective and objective image qualities of the reconstructed pictures by both systems are excellent as the 14th frame reconstruction is shown in Picture 2.1, and 2.2. The worst reconstructed images are the 6th frames shown in Picture 2.3, and 2.4. Comparing the original 6th frame and 14th frame shown in Picture 1.1, and 1.2, with the reconstructed ones, it is observed that a so-called ghost effect slightly degraded the facial part but not the background of the reconstructed image.

It is noteworthy that the frame difference sometimes contains noise in the stationary area as shown in Picture 4.1. This noise can be filtered out in order to reduce the number of moving pixels to be buffered.

Conclusions and further research

We have described an interframe data compression system which uses an adaptive strategy of replenishment suspension and 2 bits/pixel IDPCM spatial compression scheme within the framework of Hybrid optical/digital architecture. The results of this temporally adaptive interframe compression simulation are encouraging because of the overall good objective and subjective image quality (MSE = 0.08 ~ 1.67%), the average number of bits per frame: 64 Kbits/frame, the average compression ratio per frame: 16.84, and the transmission rate with frame rate 30 frames/sec: 1.92 Mbits.

There are some obvious directions in which future research in this scheme should be pointed. For example, the motion detection scheme can be modified by filtering the low

292-10

contrast moving areas and the frame-to-frame noise in the static areas of the picture, or by an adaptive reduction in the sensitivity of the motion detector. The next step of raising the efficiency and adaptability of the interframe compression system seems to be the development of the more efficient and real-time motion detection processor within the frame work of hybrid optical/digital architecture.

Also, it is worth examining the minimum size of the frame buffer in order that the subjective image quality of the image reconstructed by the nonadaptive conditional replenishment coding is sufficiently good. Usually, the buffer size is constrained by the progressing digital memory technology and cost so that it may cease to be the major limiting factor to design the interframe coder.

Finally, it is important to incorporate the effect of address coding and channel coding under the channel error.

Acknowledgements

This research was supported by the U. S. Air Force under Grant No. AFOSR-76-3024

References

1. B. R. Hunt, and H. N. Ito, "A Hybrid Optical/Digital Interframe Image Data Compression Scheme", SPIE Proceedings, Vol. 249, 1980.
2. Yasuda, H. F. Kanaya, and H. Kawawishi, "1.544 bits/s transmission of TV signals by interframe coding system", IEEE Trans. Comm., Vol. COM-24, pp. 1175-1180, Oct. 1976.
3. H. N. Ito, Image Coding Systems for Image Data Compression: A Survey, Digital Image Analysis Laboratory Report, Department of Systems Engineering, The University of Arizona, Tucson, Arizona, 85721, September, 1979.
4. A. N. Netravali, and J. O. Limb, "Picture Coding. A Review", Proceedings of IEEE, Vol. 68, No. 3, pp. 366-406, March 1980.
5. B. R. Hunt, "Optical Computing for Image Bandwidth Compression: Analysis and Simulation", Applied Optics, Vol. 17, pp. 2944-2951, September 15, 1978.
6. B. R. Hunt, and S. D. Cabrera, "Optical Implementation of a Spatially Adaptive Image Data Compression System", SPIE Proceedings, April 7-11, 1980.
7. B. R. Hunt, and H. N. Ito, "A Hybrid Optical/Digital Interframe Image Data Compression Scheme", SPIE Proceedings, July, 1980.
8. H. N. Ito, "Hybrid Image Processing Using Opto-Electronic Devices: A Survey and Perspective for Hybrid Image Processing Systems Design", Digital Image Analysis Laboratory Report, Department of Systems Engineering, The University of Arizona, Tucson, Arizona, 85721, December, 1980.
9. C. L. Fennema, and W. B. Thompson, "Velocity Determination in Scenes Containing Several Moving Objects", Computer Graphics and Image Processing, Vol. No. 9, pp. 331-335, 1979.
10. J. O. Limb, and J. A. Murphy, "Measuring the Speed of Moving Objects from Television Signals", IEEE Trans. Comm., April, 1975, pp 474-478.
11. T. S. Huang, and A. Y. Tsai, "Image Sequence Analysis: Motion Estimation," Chapter 1, of Image Sequence Analysis, ed. by T. S. Huang, Springer Verlag, 1981.
12. C. Cafforio, and P. Rocca, "Methods for Measuring Small Displacements of Television Images", IEEE Trans. Inform. Th., Vol. IT-22, pp. 573-579, September, 1976.
13. A. N. Netravali, A. N. and J. D. Robbins, "Motion Compensated Coding: Some New Results", B.S.T.J., Vol. 59, No. 9, pp. 1735-1745, November, 1980.
14. B. R. Hunt "Optical Computations for Image Bandwidth Compression", Technical Report, S2H/DIAL-81-005, Digital Image Analysis Laboratory, Department of Systems Engineering, University of Arizona, Tucson, Arizona, 85721

Copy available to DTIC does not
permit fully legible reproduction

292-10

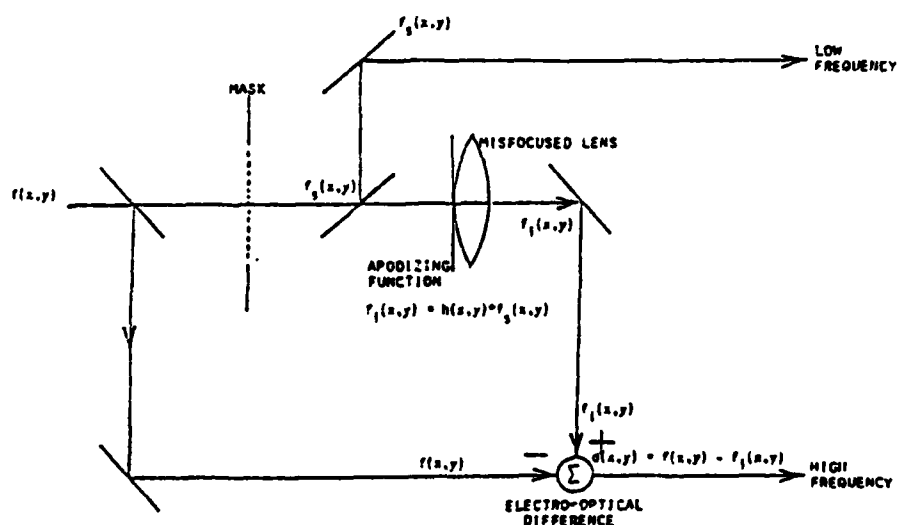


FIGURE 1(a) INCOHERENT OPTICAL SYSTEM OF IDTCI COMPRESSION

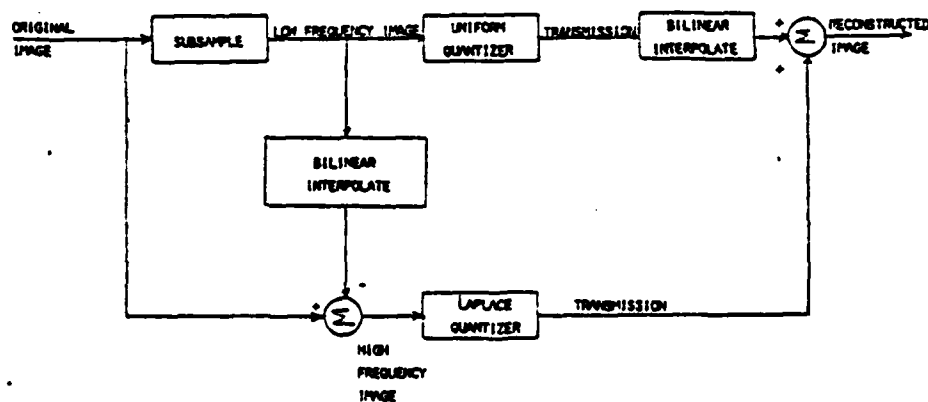


Figure 1(b) Interpolative DPCM System Block Diagram

2	4	8	16
14.37%	14.32%	14.37%	14.38%
1.0 bits/pixel	1.75 bits/pixel	2.5 bits/pixel	3.25 bits/pixel
2.816%	2.488%	2.676%	2.478%
1.25 bits/pixel	2.0 bits/pixel	2.75 bits/pixel	3.5 bits/pixel
0.837%	0.678%	0.4887%	0.6481%
1.5 bits/pixel	2.25 bits/pixel	3.0 bits/pixel	3.75 bits/pixel
0.1988%	0.1638%	0.1481%	0.1488%
1.75 bits/pixel	2.5 bits/pixel	3.25 bits/pixel	4.00 bits/pixel
0.08718%	0.08403%	0.03338%	0.02997%
2.00 bits/pixel	2.75 bits/pixel	3.5 bits/pixel	4.25 bits/pixel
0.08638%	0.0380%	0.01619%	0.01917%
2.25 bits/pixel	3.0 bits/pixel	3.75 bits/pixel	4.5 bits/pixel

Table 1(a)

Copy available to DTIC does not
contain fully legible reproduction

LPUG	NFC	2	4	8	16
2	13.525	13.075	13.045	13.045	13.045
	1.0 bits/pixel	1.5375 bits/pixel	2.875 bits/pixel	3.8125 bits/pixel	3.8125 bits/pixel
4	2.785	2.4175	2.2715	2.2385	2.2385
	1.5625 bits/pixel	2.0 bits/pixel	2.9375 bits/pixel	3.875 bits/pixel	3.875 bits/pixel
8	0.9935	0.86635	0.46985	0.41915	0.41915
	1.175 bits/pixel	2.0625 bits/pixel	3.0 bits/pixel	3.9375 bits/pixel	3.9375 bits/pixel
16	0.72295	0.37875	0.20195	0.16015	0.16015
	1.1875 bits/pixel	2.125 bits/pixel	3.0625 bits/pixel	4.0 bits/pixel	4.0 bits/pixel
32	0.6025	0.22825	0.10825	0.06094	0.06094
	1.25 bits/pixel	2.1675 bits/pixel	3.125 bits/pixel	4.0625 bits/pixel	4.0625 bits/pixel
64	0.59685	0.25965	0.092735	0.041085	0.041085
	1.3125 bits/pixel	2.25 bits/pixel	3.8125 bits/pixel	4.125 bits/pixel	4.125 bits/pixel

Table 1(b)

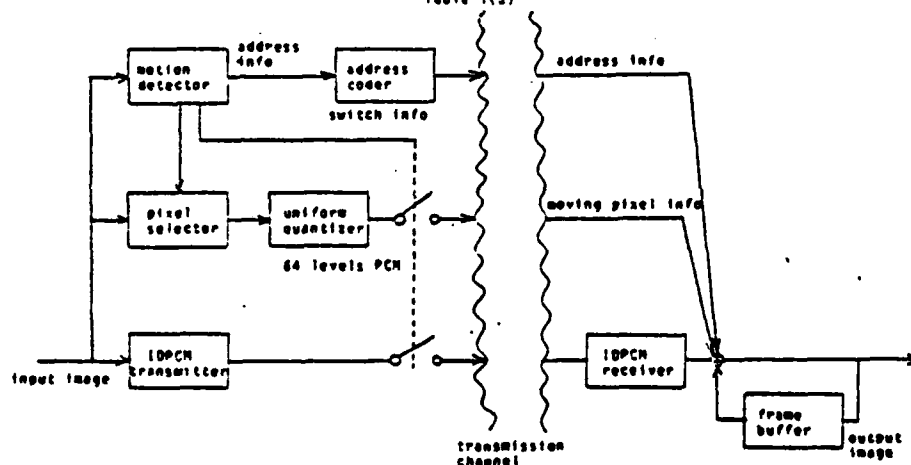


Figure 2. Hybrid Optical/Digital Adaptive (IDPCM/Conditional) Replenishment Interframe Compression Architecture

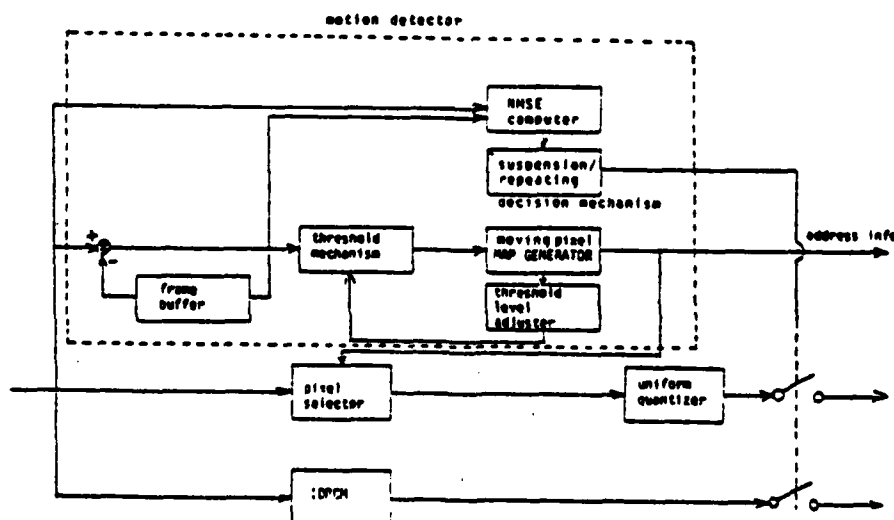
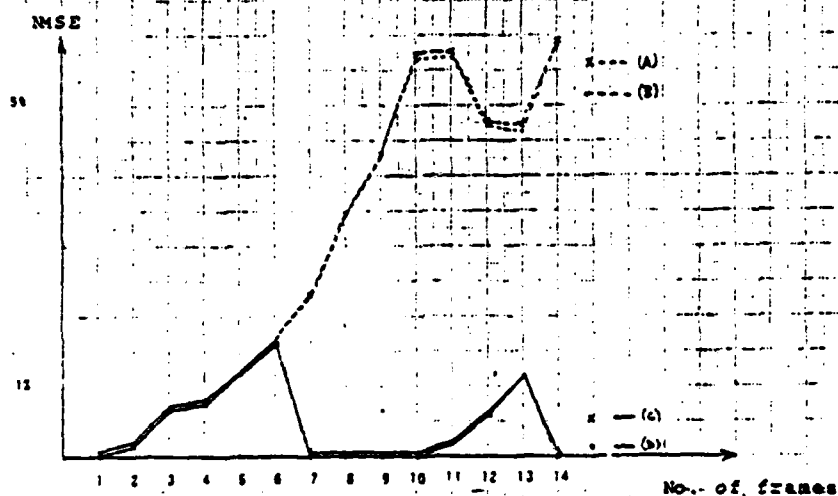


Figure 3. Simple Motion Detection Scheme

Copy available to DTIC does not permit fully legible reproduction

	NONADAPTIVE	ADAPTIVE
PCM based conditional redistribution system	(A) 41.35 kb/sec/frame 1.24 mb/sec/sec ACR = 29.96	(C) 148.34 kb/sec/frame 4.15 mb/sec/sec ACR = 16.85
DPCM based conditional redistribution system	(B) 22.263 kb/sec/frame 0.68 mb/sec/sec ACR = 26.18	(D) 54.1 kb/sec/frame 1.32 mb/sec/sec ACR = 17.06

Table 2. Transmission Bit Rate and Compression Ratio Performances of 4 systems (A), (B), (C), (D)



Graph 1. NMSE performance transition of 4 systems in 14 frames

Reconstructed Frame #	NMSE	no. of Moving Pixels Transmitted	Compression Ratio	no. of Bits Transmitted
1	0.087195	256x256	4.0	131070
2	0.23796	2071	31.648	12426
3	0.72871	2234	29.342	13404
4	0.81175	2191	29.911	13146
5	1.3245	2272	28.243	13632
6	1.6665	2152	30.410	12920
7	3.38495	256x256	1.0	131070
8	0.382725	256x256	1.0	131070
9	0.079665	256x256	1.0	131070
10	3.380485	256x256	1.0	131070
11	0.28245	2842	24.823	18882
12	3.88125	2087	32.446	12822
13	1.1215	2931	23.297	15408
14	0.034755	256x256	1.0	131070

Table 3. Performance of the adaptive system (A) (B) (C) (D) (E) (F) (G) (H) (I) (J) (K) (L) (M) (N) (O) (P) (Q) (R) (S) (T) (U) (V) (W) (X) (Y) (Z) (AA) (AB) (AC) (AD) (AE) (AF) (AG) (AH) (AI) (AJ) (AK) (AL) (AM) (AN) (AO) (AP) (AQ) (AR) (AS) (AT) (AU) (AV) (AW) (AX) (AY) (AZ) (BA) (BB) (BC) (BD) (BE) (BF) (BG) (BH) (BI) (BJ) (BK) (BL) (BM) (BN) (BO) (BP) (BQ) (BR) (BS) (BT) (BU) (BV) (BW) (BX) (BY) (BZ) (CA) (CB) (CC) (CD) (CE) (CF) (CG) (CH) (CI) (CJ) (CK) (CL) (CM) (CN) (CO) (CP) (CQ) (CR) (CS) (CT) (CU) (CV) (CW) (CX) (CY) (CZ) (DA) (DB) (DC) (DD) (DE) (DF) (DG) (DH) (DI) (DJ) (DK) (DL) (DM) (DN) (DO) (DP) (DQ) (DR) (DS) (DT) (DU) (DV) (DW) (DX) (DY) (DZ) (EA) (EB) (EC) (ED) (EE) (EF) (EG) (EH) (EI) (EJ) (EK) (EL) (EM) (EN) (EO) (EP) (EQ) (ER) (ES) (ET) (EU) (EV) (EW) (EX) (EY) (EZ) (FA) (FB) (FC) (FD) (FE) (FF) (FG) (FH) (FI) (FJ) (FK) (FL) (FM) (FN) (FO) (FP) (FQ) (FR) (FS) (FT) (FU) (FV) (FW) (FX) (FY) (FZ) (GA) (GB) (GC) (GD) (GE) (GF) (GG) (GH) (GI) (GJ) (GK) (GL) (GM) (GN) (GO) (GP) (GQ) (GR) (GS) (GT) (GU) (GV) (GW) (GX) (GY) (GZ) (HA) (HB) (HC) (HD) (HE) (HF) (HG) (HH) (HI) (HJ) (HK) (HL) (HM) (HN) (HO) (HP) (HQ) (HR) (HS) (HT) (HU) (HV) (HW) (HX) (HY) (HZ) (IA) (IB) (IC) (ID) (IE) (IF) (IG) (IH) (II) (IJ) (IK) (IL) (IM) (IN) (IO) (IP) (IQ) (IR) (IS) (IT) (IU) (IV) (IW) (IX) (IY) (IZ) (JA) (JB) (JC) (JD) (JE) (JF) (JG) (JH) (JI) (JJ) (JK) (JL) (JM) (JN) (JO) (JP) (JQ) (JR) (JS) (JT) (JU) (JV) (JW) (JX) (JY) (JZ) (KA) (KB) (KC) (KD) (KE) (KF) (KG) (KH) (KI) (KJ) (KK) (KL) (KM) (KN) (KO) (KP) (KQ) (KR) (KS) (KT) (KU) (KV) (KW) (KX) (KY) (KZ) (LA) (LB) (LC) (LD) (LE) (LF) (LG) (LH) (LI) (LJ) (LK) (LL) (LM) (LN) (LO) (LP) (LQ) (LR) (LS) (LT) (LU) (LV) (LW) (LX) (LY) (LZ) (MA) (MB) (MC) (MD) (ME) (MF) (MG) (MH) (MI) (MJ) (MK) (ML) (MM) (MN) (MO) (MP) (MQ) (MR) (MS) (MT) (MU) (MV) (MW) (MX) (MY) (MZ) (NA) (NB) (NC) (ND) (NE) (NF) (NG) (NH) (NI) (NJ) (NK) (NL) (NM) (NN) (NO) (NP) (NQ) (NR) (NS) (NT) (NU) (NV) (NW) (NX) (NY) (NZ) (OA) (OB) (OC) (OD) (OE) (OF) (OG) (OH) (OI) (OJ) (OK) (OL) (OM) (ON) (OO) (OP) (OQ) (OR) (OS) (OT) (OU) (OV) (OW) (OX) (OY) (OZ) (PA) (PB) (PC) (PD) (PE) (PF) (PG) (PH) (PI) (PJ) (PK) (PL) (PM) (PN) (PO) (PP) (PQ) (PR) (PS) (PT) (PU) (PV) (PW) (PX) (PY) (PZ) (QA) (QB) (QC) (QD) (QE) (QF) (QG) (QH) (QI) (QJ) (QK) (QL) (QM) (QN) (QO) (QP) (QQ) (QR) (QS) (QT) (QU) (QV) (QW) (QX) (QY) (QZ) (RA) (RB) (RC) (RD) (RE) (RF) (RG) (RH) (RI) (RJ) (RK) (RL) (RM) (RN) (RO) (RP) (RQ) (RR) (RS) (RT) (RU) (RV) (RW) (RX) (RY) (RZ) (SA) (SB) (SC) (SD) (SE) (SF) (SG) (SH) (SI) (SJ) (SK) (SL) (SM) (SN) (SO) (SP) (SQ) (SR) (SS) (ST) (SU) (SV) (SW) (SX) (SY) (SZ) (TA) (TB) (TC) (TD) (TE) (TF) (TG) (TH) (TI) (TJ) (TK) (TL) (TM) (TN) (TO) (TP) (TQ) (TR) (TS) (TT) (TU) (TV) (TW) (TX) (TY) (TZ) (UA) (UB) (UC) (UD) (UE) (UF) (UG) (UH) (UI) (UJ) (UK) (UL) (UM) (UN) (UO) (UP) (UQ) (UR) (US) (UT) (UU) (UV) (UW) (UX) (UY) (UZ) (VA) (VB) (VC) (VD) (VE) (VF) (VG) (VH) (VI) (VJ) (VK) (VL) (VM) (VN) (VO) (VP) (VQ) (VR) (VS) (VT) (VU) (VV) (VW) (VX) (VY) (VZ) (WA) (WB) (WC) (WD) (WE) (WF) (WG) (WH) (WI) (WJ) (WK) (WL) (WM) (WN) (WO) (WP) (WQ) (WR) (WS) (WT) (WU) (WV) (WW) (WX) (WY) (WZ) (XA) (XB) (XC) (XD) (XE) (XF) (XG) (XH) (XI) (XJ) (XK) (XL) (XM) (XN) (XO) (XP) (XQ) (XR) (XS) (XT) (XU) (XV) (XW) (XX) (XY) (XZ) (YA) (YB) (YC) (YD) (YE) (YF) (YG) (YH) (YI) (YJ) (YK) (YL) (YM) (YN) (YO) (YP) (YQ) (YR) (YS) (YT) (YU) (YV) (YW) (YX) (YY) (YZ) (ZA) (ZB) (ZC) (ZD) (ZE) (ZF) (ZG) (ZH) (ZI) (ZJ) (ZK) (ZL) (ZM) (ZN) (ZO) (ZP) (ZQ) (ZR) (ZS) (ZT) (ZU) (ZV) (ZW) (ZX) (ZY) (ZZ)

Copy available to DTIC does not
permit fully legible reproduction



Picture 1.1
nonadaptive PCM based conditional
replenishment 14th frame
reconstruction



Picture 1.2
nonadaptive IDPCM based
conditional replenishment
14th frame reconstruction



Picture 2.1
adaptive PCM based condi-
tional replenishment 14th
frame reconstruction



Picture 2.2
adaptive IDPCM based
conditional replenishment
14th frame reconstruction



Picture 2.3
adaptive PCM based
conditional replenishment
6th frame reconstruction



Picture 2.4
adaptive IDPCM based
conditional replenishment
6th frame reconstruction



Picture 3.2
original 14th frame



Picture 3.1
original 6th frame



Picture 4.1
frame difference
between 6th and 7th frames

Appendix III

Radiometric and Spatial Transformations for Image Data Compression

Radiometric and Spatial Transformations for Image Data Compression

Research Objectives

Adaptive image data compression has been exclusively a digital processor function because of the nonlinear and space-variant operations. However, a scanning optical system can achieve geometric space-variant operations by alteration of the optical system's instantaneous field of view (IFOV) and/or variation of the path of the IFOV. Radiometric space-variance can be achieved by the usage of a programmable mask, e.g., a PROM or liquid crystal. In the research under this task we are exploring the extent to which radiometric and geometric space-variance can improve the performance of a simple non-adaptive optical compression scheme, IDPCM, previously developed under sponsorship of Grant AFOSR-76-3024.

Appendix Contents:

This appendix consists of a summary of research results on employing radiometric and spatial transforms, plus a paper on the theoretical basis for deriving spatial transformations.

1) INTRODUCTION

An optical means for intraframe compression, IDPCM, originally developed by Hunt¹, is modified by pre-compression and post-compression processing. The pre-compression processing takes the form of a radiometric and/or spatial transformation performed on the original image to produce an image with stationary mean and energy (radiometric) and/or autocorrelation (spatial), respectively, as defined over sub-blocks of the image. The post-compression processing is simply the inverse of whatever transformation that was applied to the original image.

Such a modification to the basic compression step is possibly a way to lower bit rates for given image quality, or improve image quality at a given bit rate.

The analysis that follows describes the effect of various radiometric transformations on objective and subjective image quality, for different compression bit rates. A spatial transformation is also investigated.

2) PROCEDURE

The image used in this analysis is the 128 X 128 pixel, 256 grey level (8 bit/pixel) Walter picture shown in Fig. 1. The small size is necessary to keep processing time to a minimum. A 256 X 256 8 bit/pixel image is also processed for a few of the cases described below, to determine if image size is important.

Radiometric

The general procedure is as follows; the original image is compressed with the IDPCM architecture alone for various values of low frequency quantization value L and high frequency quantization value H, to define a set of images that are purely the result of the bit compression. Then another set of images are derived for the same L,H values as in the first set, but with the further processing steps of a) performing a radiometric transformation on the image before the compression step and b) performing the inverse radiometric transformation on the compressed image from part a. The radiometric transform used is that derived by Strickland², namely:

$$g(j,k) = \frac{\sigma_s}{\sigma_n} [f(j,k) - \mu_N] + \mu_s \quad (1)$$

where $f(j,k)$ is the original value of the $j^{\text{th}}, k^{\text{th}}$ pixel,

$g(j,k)$ is the transformed value of the $j^{\text{th}}, k^{\text{th}}$ pixel,

μ_N is the original value of the mean determined over each sub-block of the image (in all transformations described below, the sub-blocks are 8 X 8 pixel squares; there are 256 of these sub-blocks in the 128 X 128 image used)

μ_s is the new value of the sub-block mean (the same for all sub-blocks)

σ_N is the original standard deviation determined over each sub-block, and

σ_s is the new value of the sub-block standard deviation (the same for all sub-blocks).

The subscripts s and N refer to stationary and non-stationary, respectively. By stationary, it is meant that the statistics (namely mean and standard deviation) of each sub-block after the above transformation will be the same as those of any other sub-block. The parameters of this transformation, μ_s and σ_s , are these stationary values; they are set by the user. (The parameter σ_s is actually determined by setting the desired stationary energy, R_s ; i.e.,

$$\sigma_s^2 = R_s - \mu_s^2 \quad (2)$$

- thus it is the energy and mean parameters that are discussed in what follows.)

This transformation is applied to each sub-block individually; a blocky image would result if no further modification was made. Therefore the coefficients μ_N and σ_N (one set for each block) are linearly interpolated between block centers to yield a smooth image. An example of this radiometric transformation acting on Fig. 1 is shown in Fig. 3b, with $\mu_s = 100$ and $R_s = 12000$. The histograms of the original and transformed images are shown in Figs. 6 and 8, respectively.

Table 1 lists the various reconstructed images. For given values of L and H the energy, R_s , of the radiometric transformation is varied to determine what effect the transformation has on the compression process. The transform/compression/inverse transform process for Fig. 3a is illustrated by the sequence of figures outlined in Table 2. The inverse transformation is simply accomplished by inverting eq. (1); i.e., solve for $f(j,k)$. Values of L from 2 to 16 are looked at to see the sensitivity of the low frequency quantization to the transformation. Note that in almost all of the transformations, the stationary mean value is 100. This value is chosen arbitrarily; the primary interest here is to study the effect of a given mean, varying the energy parameter.

The high frequency quantization value, H , was chosen arbitrarily at first to match the low frequency quantization value, L , as shown in Table 1. A subsequent analysis (see Figs. 19 and 20) is performed to isolate the effect of the high frequency value H from the compression and low frequency quantization processes, and is discussed in the results section.

Spatial

The discussion above refers to the radiometric transformation. A spatial transformation, derived by Strickland², was applied to the compression process in the same type of way as the radiometric transform; i.e., spatially transform the original image/compress the result/inverse transform the compressed image. The spatial transformation has the property of normalizing the autocorrelation width of each sub-block over the entire image, where again the sub-block is an 8×8 pixel square. The shape of the autocorrelation function for each block

is approximated by an elliptically symmetric function. The width of this function is estimated for each block in the vertical and horizontal directions; from this information each sub-block of the image is in effect expanded to varying degrees in the horizontal and vertical directions, depending on the original autocorrelation of the sub-block in the two orthogonal directions. Thus sub-blocks with high correlation (uniform grey areas) are left alone, and sub-blocks with low correlation (edges) are expanded. Fig. 16 is an example of the spatial transform used in this analysis. Fig. 17 uses the spatial transform alone with the compression step, and Fig. 18 uses both the spatial and a radiometric transformation with the compression step.

NMSE

A quantitative comparison of two images is the normalized mean square error, or NMSE. It is defined as follows:

$$NMSE = \frac{\sum_{j,k} (f_r(j,k) - f(f,k))^2}{\sum_{j,k} f(j,k)^2} \quad (3)$$

where $f_r(j,k)$ is the value of the j^{th} , k^{th} pixel of the reconstructed image, and $f(j,k)$ is that value for the original image. The NMSE values are given in Table 1 for each reconstructed image, where Fig. 1 was used as the original image.

3) RESULTS; RADIOMETRIC

L = 2, H = 2; Low frequency channel; 128 X 128 Walter image

The $L = 2, H = 2$ case will be investigated first, with regard to the radiometric transformation. Looking at the compressed image with $L = 2, H = 2$ (Fig. 2) with no radiometric transform applied, the low frequency quantization is readily apparent by the two level pixel values; bright pixels correspond to pixel values in the original image that were above some threshold, dark pixels to those that were below the threshold. Looking at Fig. 3a (radiometric transform with energy = 12000 used), a similar quantization of pixel values is obvious, again due to the threshold imposed. (The intermediate frames, Figs. 3b to 3g, used in the production of Fig. 3a will be referred to in the subsequent analysis - see Table 2.) However, in the latter image, the IDPCM induced bright and dark areas (quantization noise) are uniformly spread over the image. This is the effect of the stationary transformation on the original picture block means, and the two-level quantization. The amount of this low frequency quantization noise can be seen, looking at the progression of Figs. 4, 3a, and 5, to be controllable by adjusting the parameter of energy, R_s , from 10,500 to 12,000, to 20,000, respectively. This parameter is ultimately affecting the histogram width of the transformed image (before compression), and this width affects the position of the uniform, equally-spaced pixel bins in the uniform quantization step (low frequency channel). This may in fact be the most important effect of the energy parameter for low quantization values. For the case here, two quantization levels, a slight change in the upper or lower limits of the histogram to be quantized causes the breakpoint pixel value to change. It so happens that in this image,

the pixel values of the background wall are very close to this cutoff value, and their quantized representation is thus sensitive to particular histogram upper and lower bounds, which, as stated above, are related to the energy parameter of the transform. (The greater the energy, the greater the spread in upper and lower bounds.) Note that Fig. 4, with its different statistics, has a breakpoint value below the wall pixel values, and the wall is uniformly bright. See Fig. 7 and Fig. 8 for examples of stationary image histograms (before compression) of Figs. 4 and 3a, respectively (the original image histogram is in Fig. 6).

Another effect of the radiometric transform is visible. Recall that Fig. 3g is the reconstructed low frequency frame before the inverse transform. Applying the inverse transform to this image added to 3f produces Fig. 3a, the final frame; note the grey levels introduced into Fig. 3a. Dark areas are made brighter, and vice versa. (This is the inverse stationary process acting on the block means.)

The effects of the radiometric transformation for the case $L = 2$, $H = 2$ in the low frequency channel are now summarized:

- 1) Dark areas of the picture normally lost in the compression step are brightened (and vice versa), thus the low frequency quantization noise is redistributed more uniformly over the image.
- 2) The value of the quantization breakpoint depends upon upper and lower bounds of the histogram, which is related to the mean and energy parameters of the transformation; notice the difference between Figs. 4 and 3a.
- 3) The inverse transform re-introduces a range of grey levels that are lost in the quantization process.

L = 2, H = 2; High frequency channel

The above discussion refers to the low frequency channel; the high frequency channel is now discussed, with respect to the single frame, Fig. 3a.

Fig. 3f is the Laplace quantized Fig. 3e; these are intermediate frames that correspond to Fig. 3a. Fig. 3f is the image added to Fig. 3g, the interpolated low frequency image, to give the image which is then inverse transformed to produce Fig. 3a. Note that the quantized image is a poor representation of the high frequency image: the significant edges are present in the quantized image, but the added quantization noise is uniform over the picture, degrading the information content. Thus, in all the frames shown at $H = 2$ (L doesn't affect high frequency quantization), the high frequency channel contributes little information to the final picture. Note that in the $L = 2$, $H = 2$ sequence (and in other sequences), the high frequency noise is visible in the background.

To determine the number of quantization levels needed to render an accurate high frequency image, H was varied from 2 to 4 to 6 to 8, in the compression of an untransformed picture that was very similar to Fig. 3e. The significant results are in Figs. 19 ($H=4$), and 20 ($H=8$). The subjective fidelity of the quantized high frequency image becomes acceptable at $H = 6$ (this image, not shown, is almost identical to Fig. 20). The effect of low high frequency quantization levels should be kept in mind in the discussion of the subsequent frames (when H is less than 6).

Quantitative comparison

The normalized mean square error, or NMSE, is the quantitative

comparison of the original image with the reconstructed image. The NMSE of Fig. 2 and Fig. 4 are .131 and .069, respectively; an improvement of 50% is apparent (see Table 1). As the energy parameter increases above 10,500 (see Figs. 3a, 5), so does the NMSE. The reason the NMSE is lower in the transformed images is evident by noting that, in the untransformed image, much of the wall area of the picture has been discarded due to the low frequency quantization step; however, the transformed image has a more uniformly represented wall (depending on the particular pixel distribution that determines the quantization breakpoint). Also, the inverse transform adds grey levels to the image (corresponding to the grey levels in the original), which tends to improve the NMSE.

L = 4, H = 4

Again increasing the energy from 10,500 (Fig. 9) to 12,000 (Fig. 10) shows the alteration of the image that is characteristic of the shift in the breakpoint values. Now, however, four levels of quantization, thus, three breakpoint values, are being used in the quantization. The value of the center breakpoint is close to the single breakpoint derived for the two level quantization, thus the wall pixel pattern resembles that of Fig. 3a. Qualitatively, the quantized nature of the untransformed image, Fig. 11, has been removed in Fig. 9. Quantitatively, there is a corresponding reduction in the NMSE of the two images, from .024 to .016, respectively.

L = 8, H = 8

Here again, the quantized nature of the untransformed image, Fig. 12, is smoothed by the radiometric transform producing Fig. 13; however,

the NMSE shows virtually no change, from .004 to .003, respectively.

$L = 16, H = 16$

There is virtually no change in using the radiometric transformation for such a high low frequency quantization ($L=16$), as seen by comparing Fig. 14 (untransformed) and Fig. 15 (transformed). The NMSE is the same (.005) for both cases. The low value for $H (=2)$ indicates that all of the information present is from the low frequency channel; the high frequency channel noise can be seen superposed on the image for both cases. It is interesting to see how much information is present in the low frequency channel. Subsampling at a coarser rate may be worthwhile in reducing the bit rate for such high values of L .

256 X 256 Walter image

The 128 X 128 Walter image was used in the above sequences, to reduce computing time. Two 256 X 256 images were studied with $L = 2$, $H = 16$, to check for consistency, and gave results similar to those of the 128 X 128 images described above.

Spatial transform

Preliminary results of the spatial transform shown in Fig. 16 for the $L = 4, H = 4$ sequence are Figs. 17 and 18. Fig. 17 is the result of the spatial transform alone; Fig. 18 is a combination of the spatial and a radiometric transform. The results show little effect due to the spatial transform used; a more radical spatial transform is probably required.

4) CONCLUSIONS

- A. Radiometric transformations seem to work best (qualitatively) to improve compressed frame quality for low frequency quantizations of 4 and 8 (i.e., $L=4$, $L=8$).
- B. Breakpoint values in the uniform low frequency quantization step are sensitive to image statistics (namely, lowest and highest pixel values), and thus change with different radiometric transformations. Such changes cause dramatic (image dependent) changes in the final reconstruction; e.g., the wall pixels in the Walter image.
- C. Quantization noise introduced by the compression is more uniformly distributed over the image when the radiometric transform is applied.
- D. The inverse radiometric transform re-introduces a range of grey levels that are lost in the quantization process.
- E. The high frequency Laplace quantized image contributes little to the final reconstructed picture for H less than 6.
- F. The spatial transform (warp) tested did little to change the qualitative value of the compressed image.

5) FUTURE WORK

- A. Improve low frequency quantization algorithm; uniform quantization appears too crude. The radiometric transform may produce a histogram whose low frequency quantization can be optimized (e.g., Laplacian, Gaussian). If not, can the radiometric transform be tailored to produce an optimum histogram shape?
- B. Try subsampling at a coarser rate in the low frequency channel when the low frequency quantization number is large; say L greater than or equal to 16.
- C. Try more radical spatial transformations to provide a better test for this type of transform; also determine the effects of coarser sampling rates used in conjunction with spatial transforms.
- D. Add spatially uniform random noise to the warped image before the compression step, subtracting the exact same noise after the compression; this will reduce contouring, at the price of a granular image. Applying the inverse warp may then reduce the effect of the granularity in image areas effected most by the warp (i.e., areas of low correlation).
- E. Combine spatial and radiometric transformations.

TABLE 1 : List of Figures

<u>Figure</u>	<u>L,H</u>	<u>Mean</u>	<u>Energy</u>	<u>Description</u>	<u>NMSE</u>
1	---	----	-----	original	----
2	2,2	----	-----	IDPCM alone	.131
3a	2,2	100	12000	IDPCM, radiometric	.090
4	2,2	100	10500	IDPCM, radiometric	.069
5	2,2	100	20000	IDPCM, radiometric	.119
6	---	---	-----	histogram of Fig. 1	----
7	---	---	-----	histogram of Fig. 4	----
8	---	---	-----	histogram of Fig. 3a	----
9	4,4	100	10500	IDPCM, radiometric	.016
10	4,4	100	12000	IDPCM, radiometric	.023
11	4,4	---	-----	IDPCM alone	.024
12	8,8	---	-----	IDPCM alone	.004
13	8,8	100	11000	IDPCM, radiometric	.003
14	16,2	---	-----	IDPCM alone	.005
15	16,2	127	43000	IDPCM, radiometric	.005
16	---	---	-----	spatial only	----
17	4,4	---	-----	IDPCM, spatial	.039
18	4,4	100	10500	IDPCM, rad., spatial	.019
19	-,4	---	-----	high freq. quant.	----
20	-,8	---	-----	high freq. quant.	----

TABLE 2 : Intermediate Processing Steps of Figure 3a

<u>Step</u>	<u>Figure</u>	<u>Description</u>
1	1	original
2	3b	radiometrically transformed Fig. 1 (mean = 100, energy = 12000)
3	3c	subsampled Fig. 3b
4	3d	uniformly quantized Fig. 3c
5	3e	high frequency of Fig. 3b (derived from subtracting the linearly interpolated Fig. 3c -not illustrated- from Fig. 3b)
6	3f	Laplace quantization of Fig. 3e
7	3g	linearly interpolated Fig. 3d
8	not illustrated	the sum of Fig. 3f and Fig. 3g
9	3a	the inverse transform of the above image; the final image

6) REFERENCES

- 1) B.R. Hunt, "Optical Computing For Image Bandwidth Compression: Analysis and Simulation", Applied Optics, Vol. 17, pp. 2944-2951, 1978.
- 2) R.N. Strickland, "Transforming Images Into Statistically Stationary Behavior", submitted for publication.



Figure 1 Original Image



Figure 2 IDPCM Alone



Figure 3a IDPCM, Radiometric
mean = 100
energy = 12000



Figure 3b.

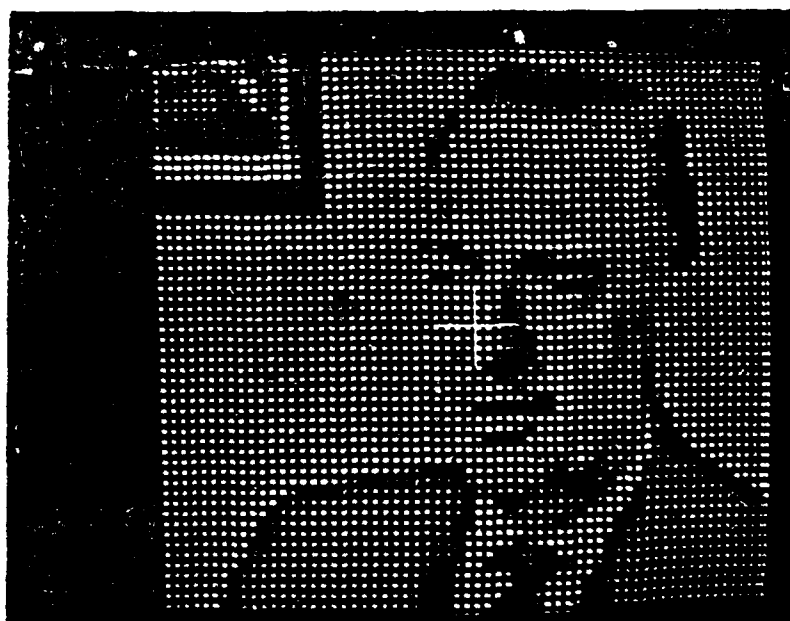


Figure 3c

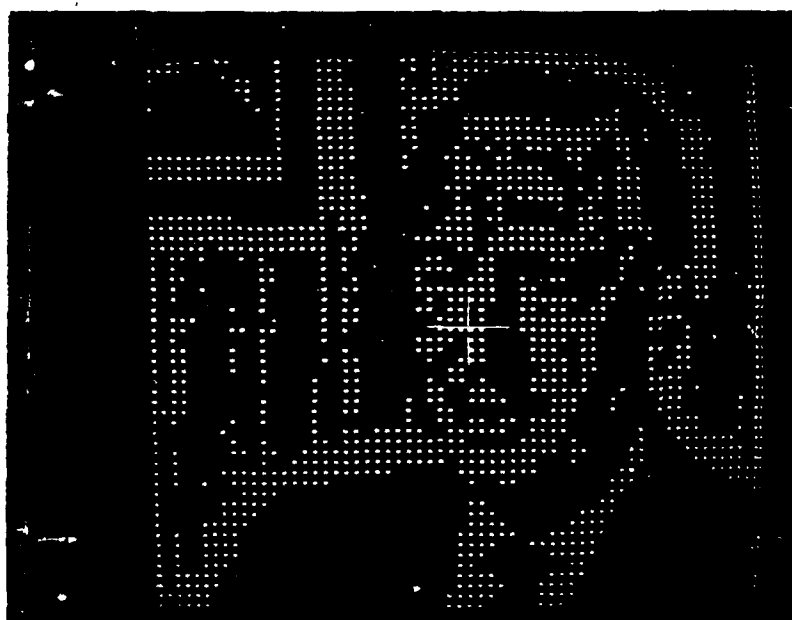


Figure 3d



Figure 3e

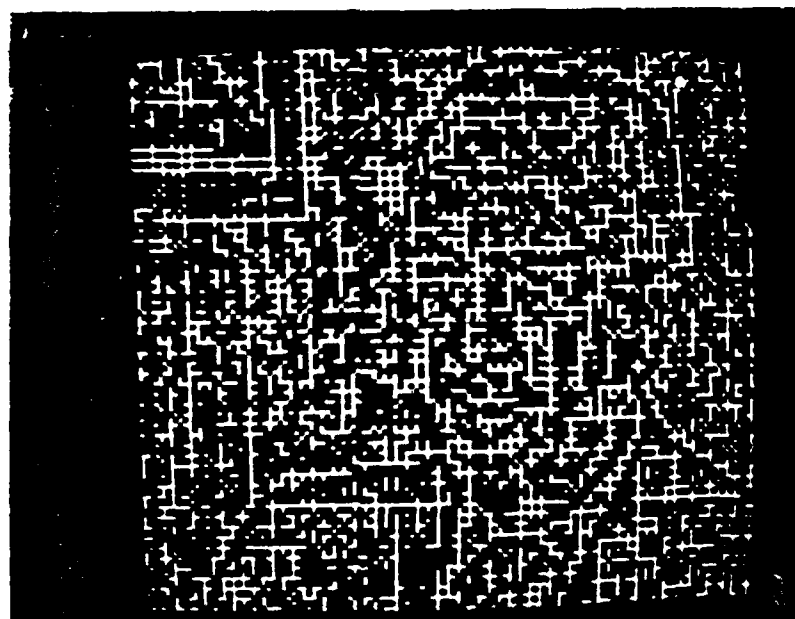


Figure 3f

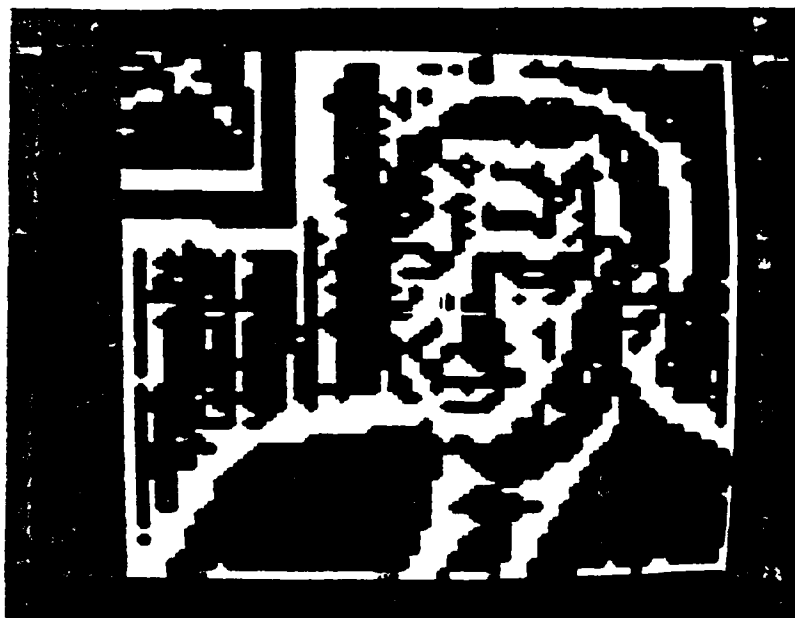


Figure 3g



Figure 4
IDPCM, Radiometric
mean = 100
energy = 10500



Figure 5
IDPCM, Radiometric
mean = 100
energy = 20000

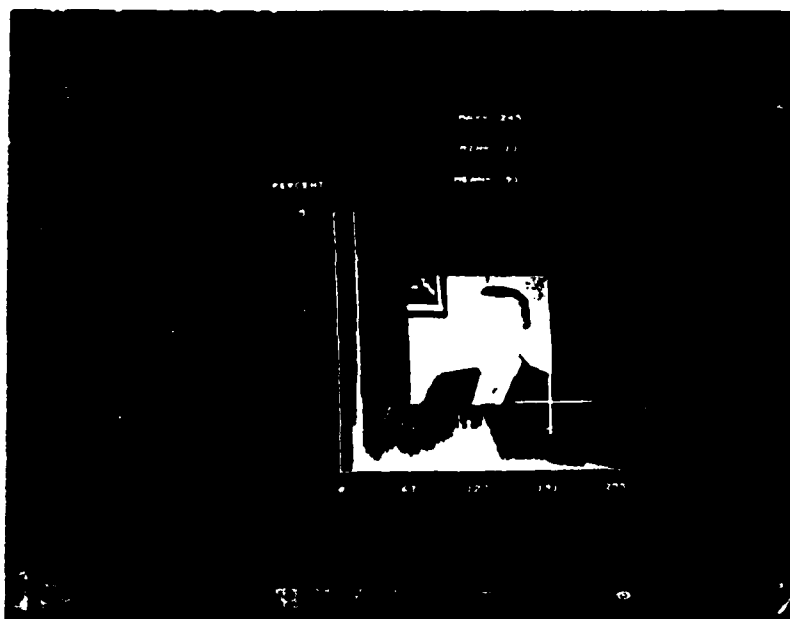


Figure 6
Histogram of Figure 1

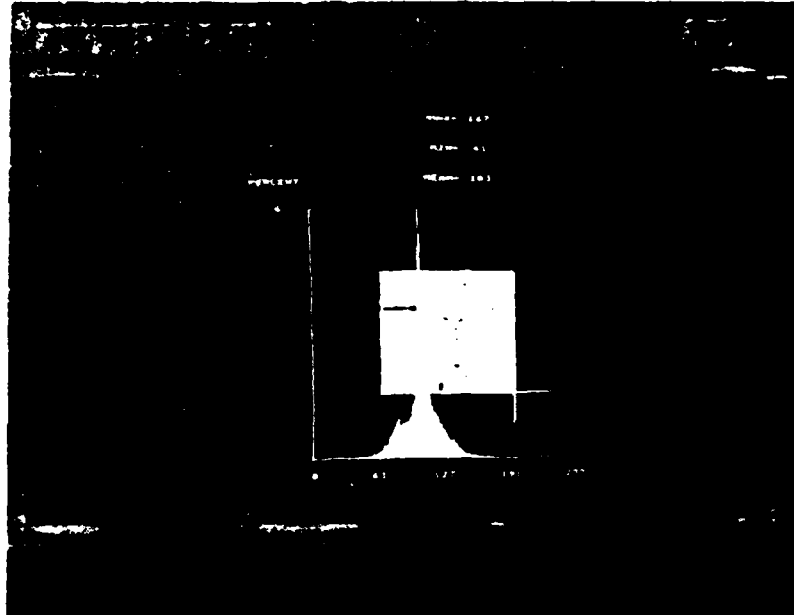


Figure 7
Histogram of Figure 4

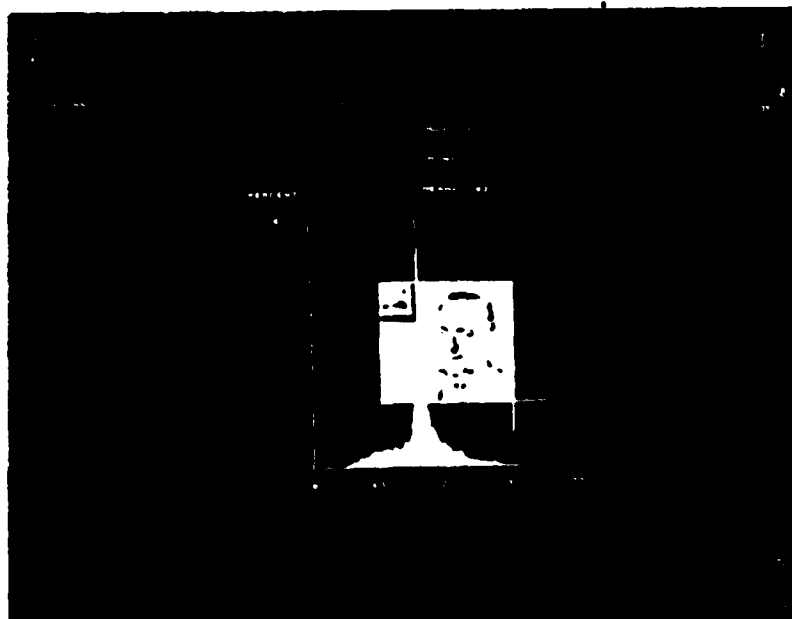


Figure 8
Histogram of Figure 3a



Figure 9
IDPCM, Radiometric
mean = 100
energy = 10500



Figure 10
IDPCM, Radiometric
mean = 100
energy = 12000



Figure 11
IDPCM Alone



Figure 12
IDPCM Alone



Figure 13
IDPCM, Radiometric
mean = 100
energy = 11000



Figure 14
IDPCM Alone



Figure 15
IDPCM, Radiometric
mean = 127
energy = 43000



Figure 16
Spatial Only



Figure 17
IDPCM, Spatial



Figure 18
IDPCM, Radiometric, Spatial
mean = 100
energy = 10500

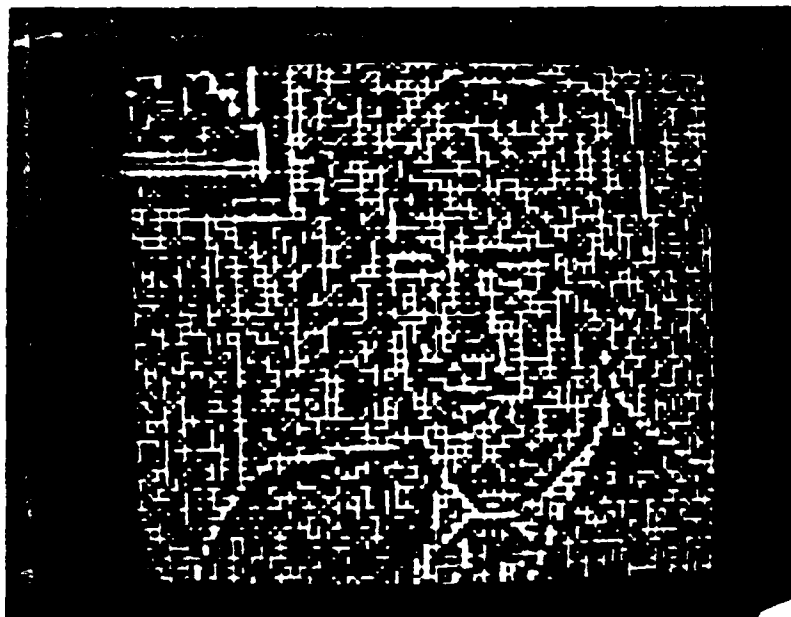


Figure 19
Laplace Quantizer with $H = 4$

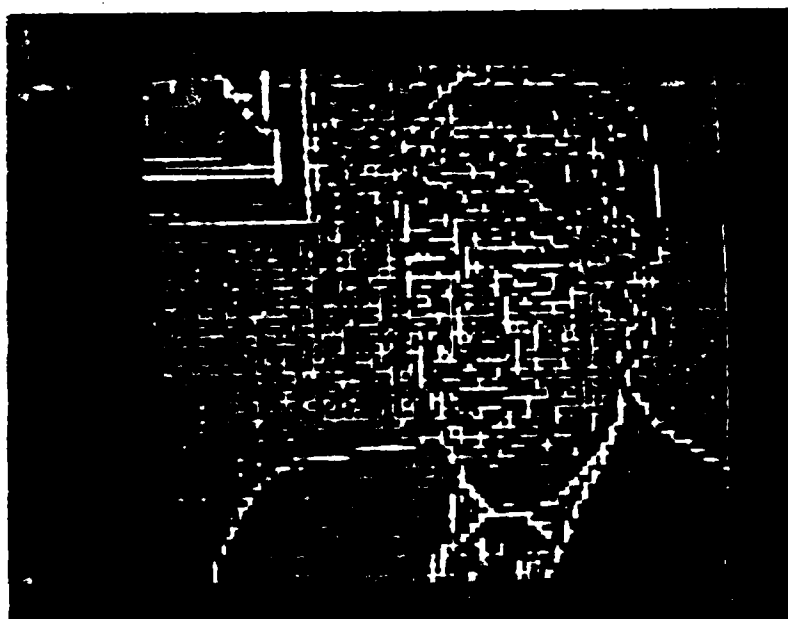


Figure 20
Laplace Quantizer with $H = 8$

ADAPTIVE DATA COMPRESSION BY TRANSFORMATIONS FOR GENERATING STATIONARY STATISTICAL IMAGE MODELS

Robin N. Strickland

Digital Image Analysis Laboratory and Optical Sciences Center, University of Arizona, Tucson, USA

1. INTRODUCTION

The statistical behavior of images is of fundamental importance in many areas of image processing. Incorporation of a priori statistical knowledge of the spatial correlation of an image, for example, can lead to considerable improvement in many image restoration algorithms. The recent trend in nonstationary restoration techniques (1),(2) has been paralleled by a similar shift in data compression schemes, from spatially nonadaptive differential pulse code modulation (DPCM) systems to adaptive DPCM systems where the coefficients of the optimum predictor are tuned to the local statistical behavior of the image. The superiority of spatially adaptive schemes is now well supported (3).

The foregoing evidence in support of using nonstationary statistical image models leads us to pose the following: suppose we turn the problem around--spatially transform or warp an image so that it possesses stationary statistics before feeding it to a nonadaptive process, such as nonadaptive DPCM or the nonadaptive Wiener filter--and then use the reverse warp transformation to restore the original image geometry. We might then expect to see improved performance from those image processes that assume stationarity.

In a previous paper, Hunt (4) proposed the use of nonstationary statistical image models in data compression. Schemes were outlined for the measurement of space-varying image parameters and implementation of image warping using hybrid digital/optical hardware. The present paper represents a continued effort toward developing transforms for generating statistically stationary images. Our emphasis will be on the application of such transforms for enhancing the performance of DPCM data compression, although we also intend to investigate their gains in image restoration applications. Throughout the paper we use digital simulation to test the transformations, although we anticipate that ultimately they can be implemented optically.

2. CONDITIONS FOR STATIONARY STATISTICS

We are concerned with generating images possessing wide-sense stationarity, involving only first- and second-order moments (5). For a two-dimensional image $f(x,y)$, these are the mean and autocorrelation statistics, respectively

$$\mu_N(x,y) = \frac{1}{N} \int_N f(x,y) dx dy \quad (1)$$

$$R_N(x,y;\xi,\eta) = \frac{1}{N} \int_N f(x,y) f(x+\xi,y+\eta) dx dy \quad (2)$$

Here the subscript N denotes that the moment statistics are measured in local neighborhoods. For the digital case, the conditions

for stationarity of the process $f(x,y)$ in the wide-sense are as follows:

$$\mu_N(j,k) = \frac{1}{M^2} \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} f(j,k) = \mu_N$$

$$= \text{constant} \quad (3)$$

and

$$R_N(j,k,r,s) = \frac{1}{M^2} \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} f(j,k) \times f(j+r,k+s) = R_N(r,s) \quad (4)$$

In other words, the means of the neighborhoods N are constant, and their autocorrelation functions have constant lag parameters. M is the dimension of a square neighborhood. In Eq. (4), an autocorrelation matrix with dimensions $(2M-1,M)$ is sufficient due to redundancy inherent in the real and even function. This property further allows us to specify the autocorrelation by four unique profiles shown in Figure 1 and designated as NS, NE, EW, and SE as points on the compass.

The spatial variance of the image statistics is reflected by the spatial variance of the mean and autocorrelation. In the case of the mean, we have a single number for each neighborhood. Somewhat more complex is the autocorrelation which, being a function of space, is characterized by its energy, width, and shape. Hence, a total of four parameters may be used to describe neighborhood statistics. According to our specifications for stationarity given in Eqs. (3) and (4), any variation between the parameters for all neighborhoods will result in nonstationary behavior.

3. THEORY OF TRANSFORMATIONS FOR STATIONARY STATISTICS

3.1. Mean and Mean-Square Energy

Consider the case of a square neighborhood N of $f(j,k)$ consisting of M^2 pixels. We assume that μ_N and $R_N(0,0)$ are initially nonstationary with respect to other neighborhoods. In general, we anticipate that stationary statistics will be generated by a simple linear transformation in intensity of the form

$$g(j,k) = Af(j,k) + B \quad (5)$$

The conditions for stationarity in $g(j,k)$ are expressed as

$$\frac{1}{M^2} \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} g(j,k) = \mu_g \quad (6a)$$

$$\frac{1}{M^2} \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} g(j,k)^2 = R_g(0,0) \quad (6b)$$

where σ_s and $R_s(0,0)$ are stationary values. Solving Eqs. (5) and (6) for constants A and B yields,

$$A = \left[\frac{\sigma_s^2 - R_s(0,0)}{\sigma_N^2 - R_N(0,0)} \right]^{1/2} = \frac{\sigma_s}{\sigma_N} \quad (7a)$$

$$= \frac{\text{stationary standard deviation}}{\text{nonstationary standard deviation}}$$

$$B = -\sigma_s - A\sigma_N \quad (7b)$$

Equations (5) and (7) combine to give the transformation for stationary mean and mean-square energy, or stationary mean and variance:

$$g(j,k) = \frac{\sigma_s}{\sigma_N} [f(j,k) - \mu_N] + \mu_s \quad (8)$$

Equation (8) represents a filtering algorithm that is identical to an algorithm developed for the spatial contrast enhancement of images (6).

In practice, Eq. (8) will generate images containing intensities that fall outside the available dynamic range of the original image, in our case 8 bits, or gray levels 0 to 255. Negative data are obviously impractical for optical implementation. Furthermore, we wish to restrict the upper gray level limit to 255 so as not to contradict our ultimate aim of data compression. We will therefore redefine the transformation in Eq. (8) to accommodate the dynamic range constraint: $0 \leq g(j,k) \leq 255$. The new transformation is

$$g(j,k) = k[f(j,k) - \mu_N] - \mu_s, \quad (9a)$$

where

$$k = \frac{\sigma_s}{\sigma_N} \text{ if } 0 \leq g(j,k) \leq 255 \quad (9b)$$

otherwise

$$k = \text{minimum} \left[\frac{255 - \mu_s}{[f(j,k) - \mu_N]_{\max}}, \frac{\mu_s}{[f(j,k) - \mu_N]_{\min}} \right] \quad (9c)$$

where $[f(j,k) - \mu_N]_{\max}$ and $[f(j,k) - \mu_N]_{\min}$ are the maximum positive-going and maximum negative-going variations (about the mean), respectively. By definition, Eq. (9) will generate images with stationary mean and approximately stationary mean-square energy.

3.2. Autocorrelation Shape

In section 2 we characterized the shape of the autocorrelation by four unique profiles. Clearly, transformation to stationary behavior requires that we assign some kind of model to each profile or, alternatively, a model for the autocorrelation as a whole. An image field is considered to be a good fit to a first-order Markov process for which the correlation between individual pixels is proportional to their geometric separation (3). We choose to neglect the diagonal profiles in Figure 1 and define an elliptically symmetric form of the autocorrelation,

$$R_{NS} = R_N(\xi, 0) = \text{constant} \cdot \exp[-\sigma_{NS}^2 \xi^2]$$

$$R_{EW} = R_N(0, \eta) = \text{constant} \cdot \exp[-\sigma_{EW}^2 \eta^2] \quad (10)$$

where R_{NS} and R_{EW} are the profiles in orthogonal North-South (vertical) and East-West (horizontal) directions respectively.

Stationary autocorrelation width along a particular direction is produced by normalizing each nonstationary σ , measured in the same direction, to some value σ_s . It is logical to equate σ_s with the minimum value of σ , corresponding to the highest correlation. The correlation widths are subsequently normalized by a resampling (interpolation) of the image data by a factor E, given by

$$E_N = \frac{\sigma_N}{\sigma_s} \quad (11)$$

E_N is therefore the expansion factor (along a given direction in the image) required to transform the autocorrelation width σ_N of a neighborhood N into a stationary value σ_s . The two-parameter model given by Eqs. (10) gives rise to expansion along orthogonal directions: E_{NS} and E_{EW} . We derive methods for performing local spatial expansions in section 5.

4. APPLYING TRANSFORMATIONS FOR STATIONARY STATISTICS

4.1. Test Image

Figure 2 shows the 256×256 pixel 8-bit "Walter" image on which we base our results. The statistics are visibly nonstationary. For example, examining contiguous blocks of 16×16 pixels, we find that the means range from 13 to 203, and the mean-squares range from 172 to 43386. The global mean is 90, with dynamic range from 9 to 249. For most of the results to follow we will consider neighborhoods of 16×16 pixels, giving a total of 256 contiguous blocks throughout the test scene. The resolution of local statistical behavior is not optimum at this block size since, ideally we would like to resolve single edges in a block.

4.2. Mean and Mean-Square Energy

We recall that the basic transformation to produce stationary mean and mean-square energy is given by Eq. (8). Applying dynamic range constraints gave a practical transform, Eq. (9). Before applying this we need to say something about the effects of noise. Let us assume for algebraic simplicity that any noise present is additive, with zero mean. We can write the noisy image as

$$f'(j,k) = f(j,k) + n(j,k), \quad (12)$$

which, when substituted for $f(j,k)$ in Eq. (8) gives

$$g(j,k) = \frac{\sigma_s}{\sigma_N} [f(j,k) + n(j,k) - \mu_N] + \mu_s \quad (13)$$

From Eq. (13) we can see that the noise term is amplified by the factor σ_s/σ_N . Ideally, we want the noise to remain in the background. In practice, we can partially attain this goal by rewriting Eq. (13) as follows:

$$g(j,k) = K[f'(j,k) - \mu_N] + \mu_s \quad (14)$$

where

$$K = \frac{\sigma_s}{\sigma_N} : \sigma_N > T$$

$$K = 1 : \sigma_N \leq T.$$

Constant T is a threshold related to the rms noise level σ_N . Assuming that $f(j,k)$ and $n(j,k)$ are uncorrelated random variables, then setting $T = \sigma_N \sqrt{2}$ would be equivalent to saying: if the rms signal variations are less or equal to the estimated noise, do not attempt to equalize the mean-square energy, or variance. Instead, adjust only the mean. The same threshold condition on K can also be applied to Eqs. (9), giving:

$$g(j,k) = K[f(j,k) - \mu_N] + \mu_s$$

where

$$K = \frac{\sigma_s}{\sigma_N} \text{ if } 0 \leq g(j,k) \leq 255,$$

otherwise

$$K = \left[\text{minimum} \frac{255 - \mu_s}{f(j,k) - \mu_N^{\max}} ; \frac{\sigma_s}{f(j,k) - \mu_N^{\min}} \right] \quad (15)$$

when $\sigma_N > T$ and

$$K = 1$$

when $\sigma_N \leq T$.

Figure 3 shows the test image after transformation using Eq. (15) with $\mu_s = 127$ and $\sigma_s = 165$, equivalent to a stationary mean-square of 43386 (as original image). We use $\sigma_N = 5$ as our noise estimate. The final image would have exhibited discontinuities at the borders between neighborhoods if we had used single values of μ_N and σ_N for each block of 16×16 pixels; instead, we use bi-linear interpolation between neighborhoods to provide unique values for every pixel. The result is a continuous image. Alternatively, we could have measured μ_N and σ_N in neighborhoods centered on all 256×256 pixels in the test image, obviously at the expense of increased computation. In Figure 3, the mean is constant at 127 (measured in 16×16 -pixel blocks), and the mean-square energy ranges from 16,000 to 26,000. We may conclude that the image is stationary in these statistics to a good approximation. The deviations in the actual mean-squares about the intended value of 43386 are due to the noise thresholding and dynamic range constraints.

4.3. Expansion Maps for Stationary Autocorrelation Shape

In section 3.2 we discussed the principles of equalizing autocorrelation shape in neighborhoods by performing localized direction-dependent expansions. After transformation to stationary mean and variance, the autocorrelation functions of contiguous 16×16 -pixel neighborhoods are computed. Least-squares fit values of σ_{NS} and σ_{EW} are computed, corresponding to the autocorrelation profiles in orthogonal directions (defined in Figure 1). The minimum values of σ for each direction from all neighborhoods are

used to generate two values of expansion for each neighborhood, as follows:

$$E_{NS} = \frac{\sigma_{NS}}{\sigma_{NS_{\min}}}, \quad E_{EW} = \frac{\sigma_{EW}}{\sigma_{EW_{\min}}} \quad (16)$$

(Note: the neighborhood subscript has been omitted from Eq. (16) for the sake of clarity.) Figures 4a and b are maps of E_{NS} and E_{EW} derived from the transformed image in Figure 3, in which each value of expansion is represented by a block of 16×16 pixels with uniform gray level. The data are scaled to 8 bits, corresponding to the range of expansion values.

5. IMAGE WARPING BASED ON EXPANSION MAPS

We now address the problem of performing geometrical warping to accommodate each one of 256 local expansions. A general approach is to assume an m -order polynomial model for the spatial distortion, giving,

$$p = \sum_{i=0}^m \sum_{j=0}^{m-1} a_{ij} x^i y^j \quad (17a)$$

$$q = \sum_{i=0}^m \sum_{j=0}^{m-1} b_{ij} x^i y^j, \quad (17b)$$

where (p,q) and (x,y) are coordinates of the two spaces (warped and unwarped, or vice versa since the transformation can take place in either direction).

A method for deriving the polynomials is provided by superposing the expansion of each neighborhood individually, with the constraint that the influence of each expansion on the relative geometry of other neighborhoods should be minimized. The technique we have developed is based on control point grids.

5.1. Control Point Mapping

Methods of using control points in similar geometrical rectification and warping applications are well documented (7,8). In our case, we define a regular grid of control points whose coordinates coincide with the corners of each square neighborhood, giving, for the previous example, a total of 17×17 control points. Our problem is to determine how this regular grid in the original image space is redistributed by the complex interactions of local expansions centered on each neighborhood. As mentioned above, we begin by determining how the grid is affected by a single expansion applied to the first neighborhood. We then apply a second expansion to the second neighborhood, and so on.

Each single expansion has two components E_{NS} and E_{EW} , acting in orthogonal directions, shown in Figure 5. The region to be expanded is contained within a circle of radius $d/\sqrt{2}$. Expansion causes vector displacements at a point (i) lying on the circumference of this circle according to the following:

$$p = \frac{d}{\sqrt{2}} \cos \theta \cdot E_{EW}, \quad q = \frac{d}{\sqrt{2}} \sin \theta \cdot E_{NS} \quad (18)$$

Outside the circle, for example at point (ii), the transformations in x and y are

$$p = \frac{d}{\sqrt{2}} \cos \theta \cdot (E_{EW} - 1) + x \quad (19a)$$

$$p = \frac{d}{2} \sin((E_{NS}-1) - \gamma). \quad (19b)$$

We see from Eqs. (18) that the four neighborhood control points lying within the circle are expanded by E_{NS} in y and E_{EW} in x . The two-dimensional transformation does not preserve the original geometry of the remaining neighborhoods. We can see intuitively that to do so is impossible. In the following section we describe a technique for reducing the mutual interference of local neighborhood expansions.

3.2. Iterative Control Point Mapping

We have extended the preceding method of successive expansions to incorporate correction factors to offset the effects of nonlocalized expansions. In Figure 6 we define the geometrical terminology for neighborhoods of arbitrary shape. The expansion components in this scheme are defined as

$$E_{NS} = \frac{c_{NS}}{c}, \quad E_{EW} = \frac{c_{EW}}{c}. \quad (20)$$

The iterative algorithm for control point mapping then proceeds as follows:

Notation:

Neighborhoods: N_1, N_2, \dots, N_{256}

Desired

expansions: $(E_{NS}, E_{EW})_1, (E_{NS}, E_{EW})_2, \dots$

Measured

expansions: $(E'_{NS}, E'_{EW})_1, (E'_{NS}, E'_{EW})_2, \dots, (E'_{NS}, E'_{EW})_{256}$

Algorithm:

Step 1

Measure $(E'_{NS}, E'_{EW})_1$ for N_1 .

Step 2

Apply corrected expansions E_{NS1}/E'_{NS1} , E_{EW1}/E'_{EW1} to N_1 .

.

.

Step 511

Measure $(E'_{NS}, E'_{EW})_{256}$ for N_{256} .

Step 512

Apply expansions E_{NS256}/E'_{NS256} , E_{EW256}/E'_{EW256} to N_{256} .

Return to Step 1, repeat for I iterations.

In the above we are applying correction factors E_{NS}, E_{EW} to the desired expansions E'_{NS}, E'_{EW} . The effectiveness of the algorithm is measured by the convergence of the rms expansion errors measured at the end of each iteration. In Figure 7 we have the warped control point map for our previous two-component expansion data after nine iterations. Figure 8 shows the converging rms error curves. There is apparently little to be gained by continuing beyond nine iterations.

3.3. Control Grid Interpolation

At this point we have two sets of control points: the first, a regular grid in original unwarped image space; the second, a severely warped grid that will ultimately establish the geometry of a stationary image. The quadrilateral vertices (control points) map directly to the corresponding square vertices. Interior points are mapped according to interpolation based upon the mapping of the vertices. We use local mapping, based on small sets of neighboring control points. For example, we can express three-

point mapping between triangles as

$$\begin{aligned} p &= a_1 + a_2x + a_3y, \\ q &= b_1 + b_2x + b_3y \end{aligned} \quad (21)$$

In Eqs. (21) we create two sets of three linear equations in the same number of unknowns by specifying that the vertices of a triangle map into the vertices of the corresponding figure in warped space. Solution of these equations leads to the mapping coefficients a and b . Global warping based on two sets of 17×17 control points is therefore achieved by computing 32×32 polynomial transformations that are applied individually to corresponding neighborhoods.

Although up until now we have talked in terms of (p, q) being warped space and (x, y) being unwarped space, in actual fact we derive the transformations after reversing the definitions. Hence, for each integral coordinate in (x, y) in output warped space, the transformations tell us where to look in unwarped (p, q) space. If, as is the general case, the coordinates (p, q) are nonintegral, then we have to resort to gray level interpolation to generate an output pixel. Bilinear interpolation of the four nearest neighbors is satisfactory. By defining the direction of the transformation from warped to unwarped space, we make the task of pixel interpolation easier.

Figure 9 shows the warped image derived by three-point spatial transformation of the intermediate image in Figure 3, based on the control point map in Figure 7. It is interesting to note that noise-smoothing by low-pass filtering would now cause relatively less blurring of visually important edges due to the spatial expansion of such regions. This is obviously a simple example of the advantages of processing stationary images.

6. CONCLUSIONS

We have discussed transformations for producing images with wide-sense stationary first- and second-order moment statistics. These transforms are reasonably straightforward to derive and to apply. We note that the inverse spatial transformation is simple to accomplish since it involves only switching the roles of the two sets of control points when implementing Eq. (21). Optical implementation of forward and reverse spatial transformations in intensity in Eq. (15) by optical means is less obvious; it may be necessary to resort to hybrid digital/optical technology. Further work will be directed at testing the effectiveness of stationary images in DPCM data compression, compared with nonstationary images.

ACKNOWLEDGMENTS

This work was performed under the sponsorship of the U.S. Air Force Office of Scientific Research under grant AFOSR-76-3024.

REFERENCES

- Anderson, G. L., and Netravali, A. N., 1976, "Image restoration based on a subjective criterion," *IEEE Trans. Sys. Man Cybern.*, 6, 845-853.
- Trussel, H. J., and Hunt, B. R., 1978, "Sectioned methods for image restoration," *IEEE Trans. Acoustics Speech Signal*

- Processing, 26, 157-164.
3. Pratt, W. K., 1978, Digital Image Processing, Wiley, New York, p. 729.
4. Hunt, B. R., 1980, "Nonstationary statistical image models (and their application to image data compression)," Computer Graphics and Image Processing, 12, 173-186.
5. Papoulis, A., 1965, Probability, Random Variables and Stochastic Processes, McGraw-Hill, New York, p. 302.
6. Wallis, R., 1976, "An approach to the space-variant restoration and enhancement of images," Proc. Symposium on Current Mathematical Problems in Image Science, Naval Postgraduate School, Monterey, California, November.
7. Bernstein, R., 1976, "Digital image processing of earth observation sensor data," IBM J. Res. Devel., 20, 40-57.
8. Strickland, R. N., and Burke, J. J., 1980, "Geometrical rectification of spin-scan images from Pioneer 11, advances in image transmission II," Proc. SPIE, 249, 47-52.



Figure 3 Image after transformation (Eq. (15)) to wide-sense stationary mean and mean-square energy.

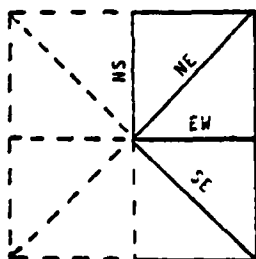
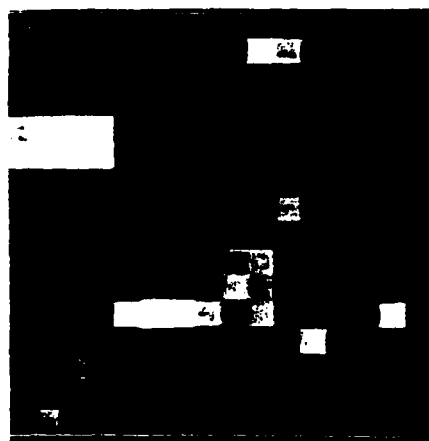


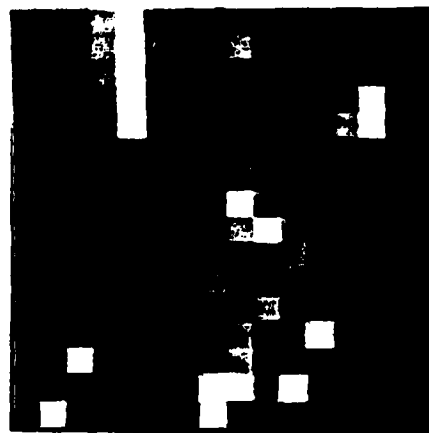
Figure 1 Autocorrelation function profile designations.



Figure 2 Original test image.



(a) N-S data



(b) E-W data

Figure 4 Expansion ratios map: elliptically-symmetric autocorrelation model

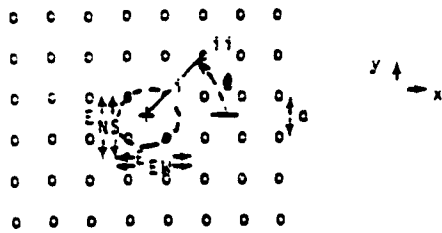
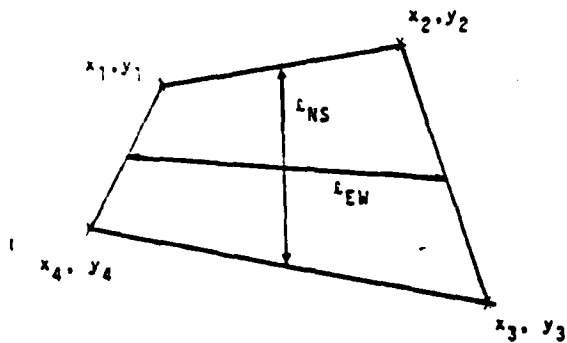


Figure 5 Principles of two-dimensional control-point mapping.



X - control points

Figure 6 Quadrilateral neighborhood geometry

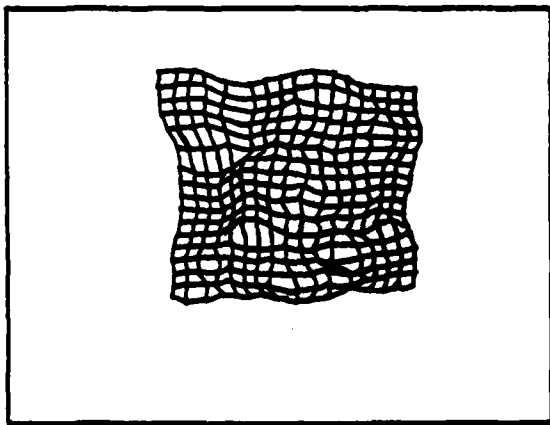


Figure 7 Iterative warping: control point grid after nine iterations (expansions of Figure 5)

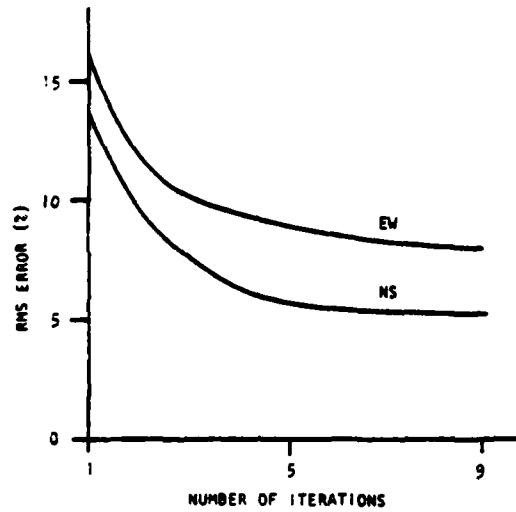


Figure 8 Convergence of iterative warping (expansions of Figure 5). (Errors are rms of 256 values.)



Figure 9. Image with wide-sense stationary mean and autocorrelation

DATE
FILMED
8-8